THE POSSIBILITIES ARE INFINITE

FUJITSU

WHITE PAPER

Version 5.1
November 2008

Performance Report PRIMERGY TX150 S6

Pages 33

Abstract

This document contains a summary of the benchmarks executed for the PRIMERGY TX150 S6.

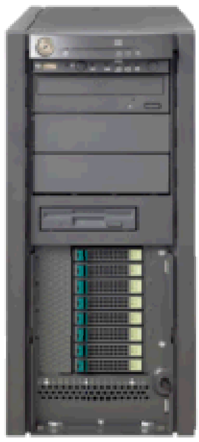
The PRIMERGY TX150 S6 performance data are compared with the data of other PRIMERGY models and discussed. In addition to the benchmark results, an explanation has been included for each benchmark and for the benchmark environment.

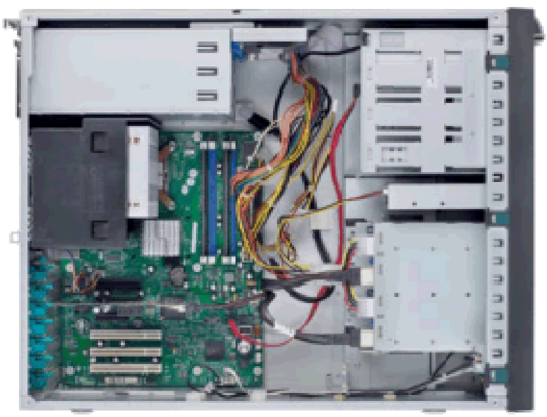
Contents

Technical Data

The PRIMERGY TX150 S6 is a mono -socket tower server. It includes the Intel 3210 chipset, one Celeron, Pentium Dual-Core, Core 2 Duo, Dual-Core Xeon or Quad-Core Xeon processor, up to 8 GB PC2-6400 DDR2-SDRAM, depending on the processor used a 800 MHz, 1067 MHz or 1333 MHz front-side bus, a Broadcom BCM5755 1-GBit LAN controller, six PCI slots (2 x PCI-Express x8, 1 x PCI-Express x4, 3 x PCI 32-bit/33 MHz) and space for four 3.5" SAS or SATA hard disks or up to 10 2.5" SAS hard disks. The SAS version of the PRIMERGY TX150 S5 has an 8-port SAS controller with RAID 0, 1 and RAID-1E functionality or - alternatively - an 8-port SAS controller with RAID 0, 1, 10, 5, 50, 6 and RAID-60 functionality. The SATA version has a 6-port SATA controller with RAID 0, 1, 10 and optionally RAID 5 functionality.

As well as its predecessors the PRIMERGY TX150 S6 can be converted quickly and easily into a rack system with 5 height units for integration in 19-inch racks.

See http://docs.ts.fujitsu.com/dl.aspx?id=af1d8493-dfbb-478e-89c9-92e917929234 for detailed technical information.

SPECcpu2006*

Benchmark description

SPECcpu2006 is a benchmark to measure system efficiency during integer and floating point operations. It consists of an integer test suite containing 12 applications and a floating point test suite containing 17 applications which are extremely computing-intensive and concentrate on the CPU and memory. Other components, such as disk I/O and network, are not measured by this benchmark.

SPECcpu2006 is not bound to a specific operating system. The benchmark is available as source code and is compiled before the actual benchmark. Therefore, the compiler version used and its optimization settings have an influence on the measurement result.

SPECcpu2006 contains two different methods of performance measurement: The first method (SPECint2006 and SPECfp2006) determines the time required to complete a single task. The second method (SPECint_rate2006 and SPECfp_rate2006) determines the throughput, i.e. how many tasks can be completed in parallel. Both methods are additionally subdivided into two measuring runs, "base" and "peak", which differ in the way the compiler optimization is used. The "base" values are always used when results are published, the "peak" values are optional.

Benchmark	Arithmetic	Type	Compiler optimization	Measuring result	Application
SPECint2006	integer	peak	aggressive	speed	single threaded
SPECint_base2006	integer	base	conservative		
SPECint_rate2006	integer	peak	aggressive	throughput	multithreaded
SPECint_rate_base2006	integer	base	conservative		
SPECfp2006	floating point	peak	aggressive	speed	single threaded
SPECfp_base2006	floating point	base	conservative		
SPECfp_rate2006	floating point	peak	aggressive	throughput	multithreaded
SPECfp_rate_base2006	floating point	base	conservative		

The results represent the geometric mean of normalized ratios determined for the individual benchmarks. Compared with the arithmetic mean, the geometric mean results in the event of differingly high single results in a weighting in favor of the lower single results. “Normalized“ means measuring how fast the test system runs in comparison to a reference system. The value of “1“ was determined for the SPECint_base2006, SPECint_rate_base2006, SPECfp_base2006 and SPECfp_rate_base2006 results of the reference system. Thus a SPECint_base2006 value of 2 means for example that the measuring system has executed this benchmark approximately twice as fast as the reference system. A SPECfp_rate_base2006 value of 4 means that the measuring system has executed this benchmark about 4/[# base copies] times as fast as the reference system. “# base copies“ here specifies how many parallel instances of the benchmark have been executed.

We do not submit all SPECcpu2006 measurements for publication at SPEC. So not all results appear on SPEC’s web sites. As we archive the log data for all measurements, we are able to prove the correct realization of the measurements any time.

Benchmark results

The PRIMERGY TX150 S6 was measured with eight different processor versions:

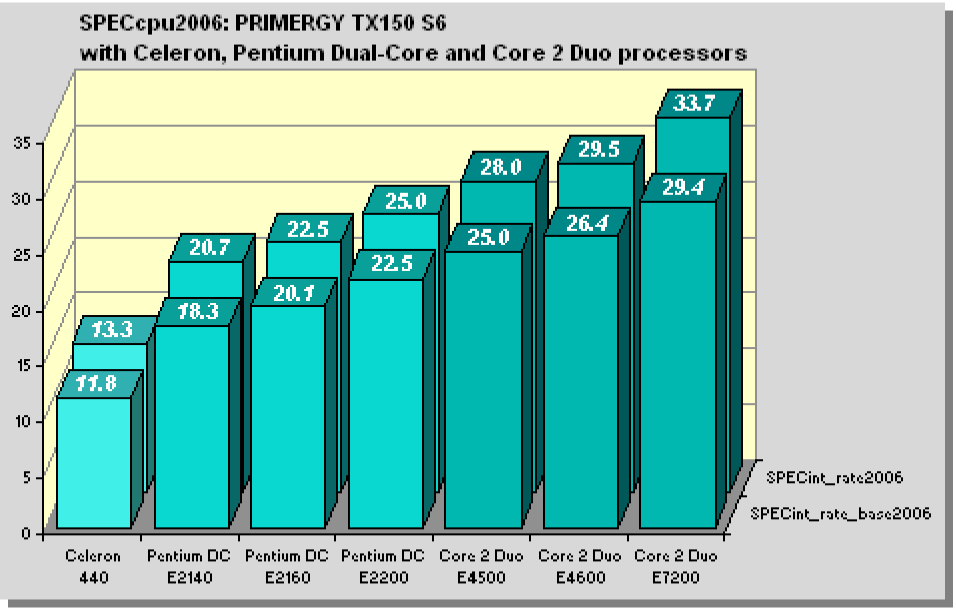
- Celeron 440
  (Conroe-L, 1 core per chip, ½ MB L2 cache per chip)
- Pentium Dual-Core E2140, E2160 and E2200
  (Allendale, 2 cores per chip, 1 MB L2 cache per chip)
- Core 2 Duo E4500 and E4600
  (Allendale, 2 cores per chip, 2 MB L2 cache per chip)
- Core 2 Duo E7200
  (Wolfdale, 2 cores per chip, 3 MB L2 cache per chip)
- Xeon 3065
  (Conroe, 2 cores per chip, 4 MB L2 cache per chip)

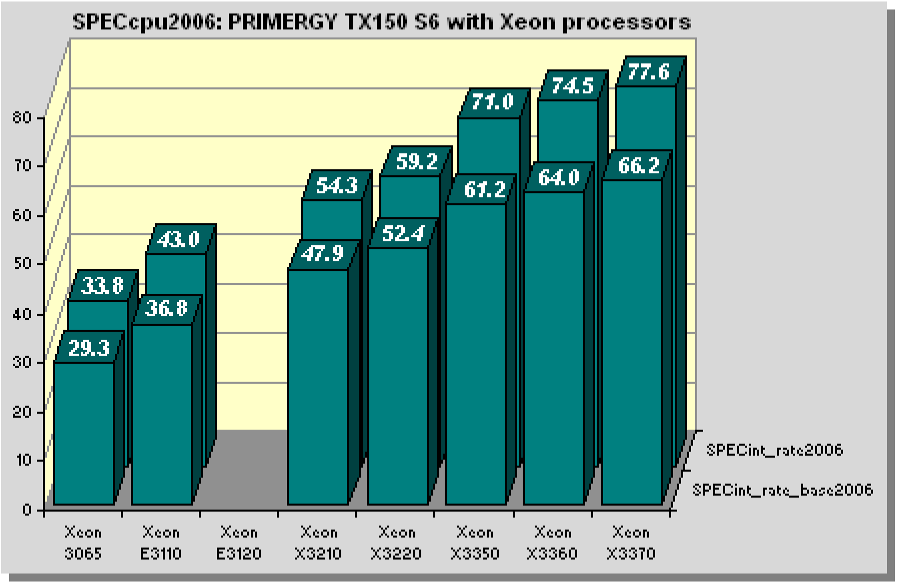
* SPEC®, SPECint®, SPECfp® and the SPEC logo are registered trademarks of the Standard Performance Evaluation Corporation (SPEC).

- Xeon E3110 and E3120
  (Wolfdale, 2 cores per chip, 6 MB L2 cache per chip)
- Xeon X3210 and X3220
  (Kentsfield, 4 cores per chip, 4 MB L2 cache per 2 cores)
- Xeon X3350, X3360 and X3370
  (Yorkfield, 4 cores per chip, 6 MB L2 cache per 2 cores)

The following two tables show results, in which all benchmark programs were compiled with the Intel C++/Fortran compiler 10.1. See the Benchmark environment section fort he operating system versions used. Bold result numbers are published at http://www.spec.org.

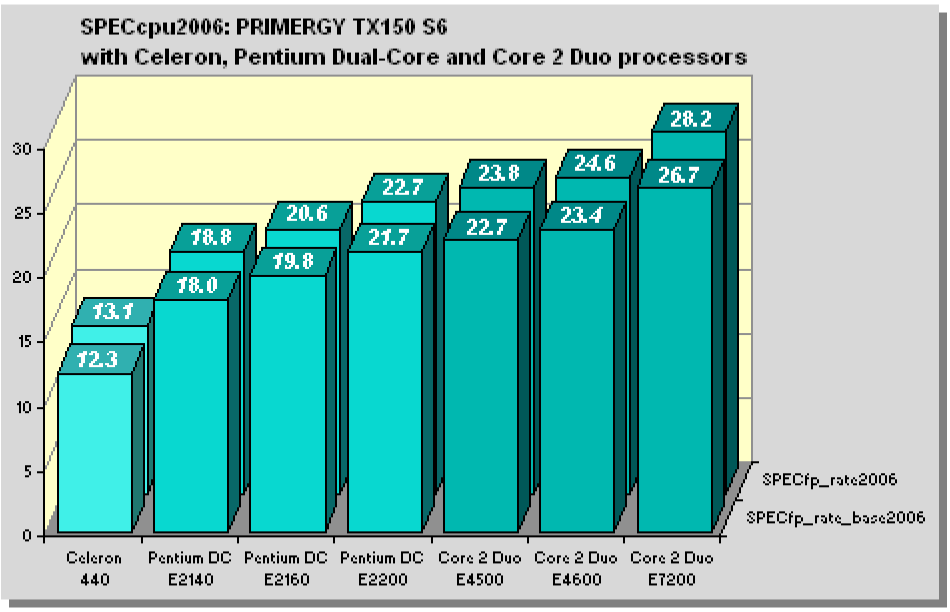
Processor	cores/ chip	GHz	L2 cache	FSB	TDP	SPECint_rate_base2006	SPECint_rate2006
Celeron 440	1	2	½ MB per chip	800 MHz	35 watt	11.8	13.3
Pentium Dual-Core E2140	2	1.60	1 MB per chip	800 MHz	65 watt	**18.3**	**20.7**
Pentium Dual-Core E2160	2	1.80	1 MB per chip	800 MHz	65 watt	**20.1**	**22.5**
Pentium Dual-Core E2200	2	2.20	1 MB per chip	800 MHz	65 watt	22.5	25.0
Core 2 Duo E4500	2	2.20	2 MB per chip	800 MHz	65 watt	**25.0**	**28.0**
Core 2 Duo E4600	2	2.40	2 MB per chip	800 MHz	65 watt	**26.4**	**29.5**
Core 2 Duo E7200	2	2.53	3 MB per chip	1067 MHz	65 watt	**29.4**	**33.7**
Xeon 3065	2	2.33	4 MB per chip	1333 MHz	65 watt	**29.3**	**33.8**
Xeon E3110	2	3	6 MB per chip	1333 MHz	65 watt	**36.8**	**43.0**
Xeon E3120	2	3.17	6 MB per chip	1333 MHz	65 watt	n/a	n/a
Xeon X3210	4	2.13	4 MB per 2 cores	1067 MHz	95 watt	**47.9**	**54.3**
Xeon X3220	4	2.40	4 MB per 2 cores	1067 MHz	95 watt	**52.4**	**59.2**
Xeon X3350	4	2.67	6 MB per 2 cores	1333 MHz	95 watt	**61.2**	**71.0**
Xeon X3360	4	2.83	6 MB per 2 cores	1333 MHz	95 watt	**64.0**	**74.5**
Xeon X3370	4	3	6 MB per 2 cores	1333 MHz	95 watt	**66.2**	**77.6**

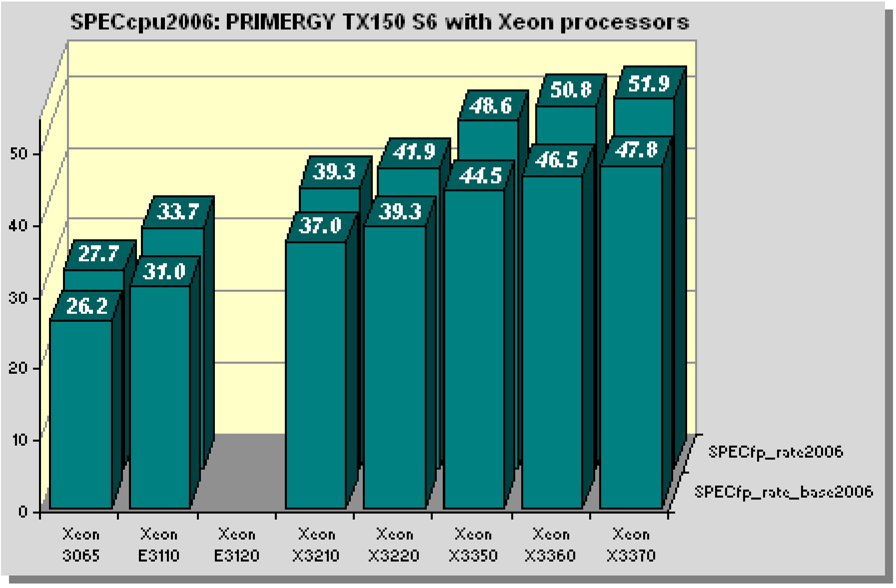




The measured SPECint_rate_2006 results are 12-17% above the SPECint_rate_base2006 results.

Processor	cores/ chip	GHz	L2 cache	FSB	TDP	SPECfp_rate_base2006	SPECfp_rate2006
Celeron 440	1	2	½ MB per chip	800 MHz	35 watt	12.3	13.1
Pentium Dual-Core E2140	2	1.60	1 MB per chip	800 MHz	65 watt	**18.0**	**18.8**
Pentium Dual-Core E2160	2	1.80	1 MB per chip	800 MHz	65 watt	**19.8**	**20.6**
Pentium Dual-Core E2200	2	2.20	1 MB per chip	800 MHz	65 watt	21.7	22.7
Core 2 Duo E4500	2	2.20	2 MB per chip	800 MHz	65 watt	**22.7**	**23.8**
Core 2 Duo E4600	2	2.40	2 MB per chip	800 MHz	65 watt	**23.4**	**24.6**
Core 2 Duo E7200	2	2.53	3 MB per chip	1067 MHz	65 watt	**26.7**	**28.2**
Xeon 3065	2	2.33	4 MB per chip	1333 MHz	65 watt	**26.2**	**27.7**
Xeon E3110	2	3	6 MB per chip	1333 MHz	65 watt	**31.0**	**33.7**
Xeon E3120	2	3.17	6 MB per chip	1333 MHz	65 watt	n/a	n/a
Xeon X3210	4	2.13	4 MB per 2 cores	1067 MHz	95 watt	**37.0**	**39.3**
Xeon X3220	4	2.40	4 MB per 2 cores	1067 MHz	95 watt	**39.3**	**41.9**
Xeon X3350	4	2.67	6 MB per 2 cores	1333 MHz	95 watt	**44.5**	**48.6**
Xeon X3360	4	2.83	6 MB per 2 cores	1333 MHz	95 watt	**46.5**	**50.8**
Xeon X3370	4	3	6 MB per 2 cores	1333 MHz	95 watt	**47.8**	**51.9**





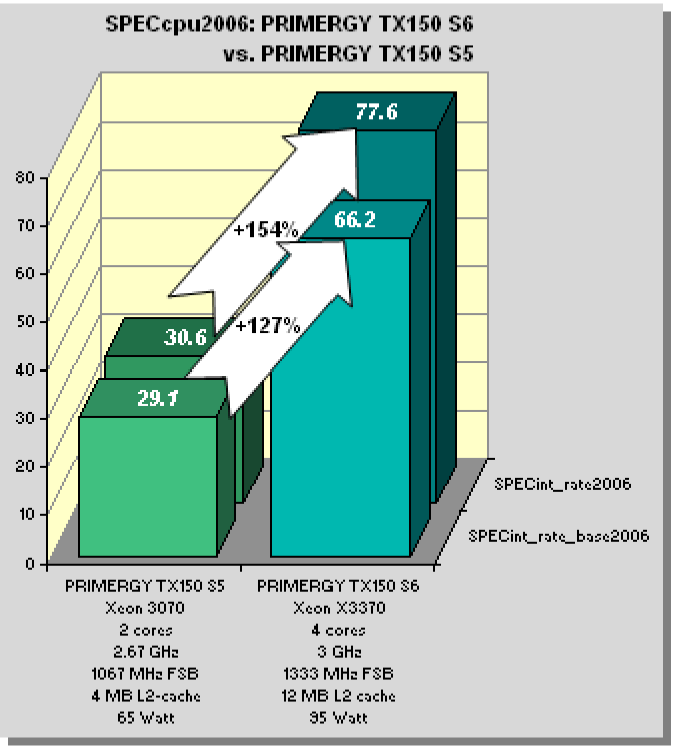
The measured SPECfp_rate_2006 results are 4-9% above the SPECfp_rate_base2006 results.

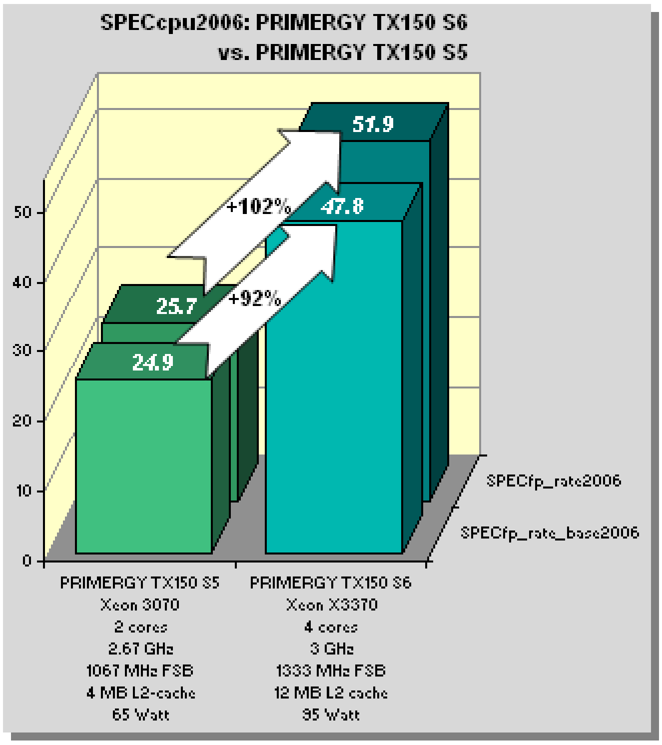
The following four tables show results, in which all benchmark programs were compiled with the Intel C++/Fortran compiler 10.1 and run under SUSE Linux Enterprise Server 10 (64-bit). Significantly higher results were achieved with the Intel C++/Fortran compiler 11.0 than would have been expected with version 10.1. In principle, this should also apply for other processors of the PRIMERGY TX150 S6.

Processor	Cores/ Chip	GHz	L2 cache	FSB	TDP	SPECint_rate_base2006	SPECint_rate2006
Xeon E3120	2	3.17	6 MB per chip	1333 MHz	65 watt	45.3	48.8

Processor	Cores/ Chip	GHz	L2 cache	FSB	TDP	SPECfp_rate_base2006	SPECfp_rate2006
Xeon E3120	2	3.17	6 MB per chip	1333 MHz	65 watt	34.3	36.3

When comparing the PRIMERGY TX150 S6 and its predecessor the PRIMERGY TX150 S5 both in their highest performance configurations, an increase is noted in the integer test suite of +127% with SPECint_rate_base2006 and +154% with SPECint_rate2006. In the floating point test suite the increase is +92% with SPECfp_rate_base2006 and +102% with SPECfp_rate2006.





Benchmark environment

All SPECcpu2006 measurements were performed on a PRIMERGY TX150 S6 with the following hardware and software configuration:

Hardware	
Model	PRIMERGY TX150 S6
CPU	Celeron 440 Pentium Dual-Core E2140, E2160 and E2200 Core 2 Duo E4500, E4600 and E7200 Xeon 3065, E3110, E3120, X3210, X3220, X3350, X3360 and X3370
Number of CPUs	1
Primary Cache	32 kB instruction + 32 kB data on chip, per core
Secondary Cache	Celeron 440: ½ MB (I+D) on chip, per chip Pentium Dual-Core E2140, E2160 and E2200: 1 MB (I+D) on chip, per chip Core 2 Duo E4500 and E4600: 2 MB (I+D) on chip, per chip Core 2 Duo E7200: 3 MB (I+D) on chip, per chip Xeon 3065: 4 MB (I+D) on chip, per chip Xeon E3110 and E3120: 6 MB (I+D) on chip, per chip Xeon X3210 and X3220: 8 MB (I+D) on chip, per chip Xeon X3350, X3360 and X3370: 12 MB (I+D) on chip, per chip
Memory	8 GB PC2-6400 DDR2-SDRAM
Software	
Operating System	Xeon 3065: SUSE Linux Enterprise Server 10 (64-bit) Pentium Dual-Core E2200, Core 2 Duo E7200, Xeon E3120 and Xeon X3370: SUSE Linux Enterprise Server 10 SP2 (64-bit) others: SUSE Linux Enterprise Server 10 SP1 (64-bit)
Compiler	Xeon E3120: Intel C++/Fortran Compiler 11.0 others: Intel C++/Fortran Compiler 10.1

SPECjbb2005*

Benchmark description

SPECjbb2005 is a Java business benchmark that focuses on the performance of Java server platforms. It is essentially a modernized version of SPECjbb2000 with the main differences being:

- The transactions have become more complex in order to cover a greater functional scope.
- The working set of the benchmark has been enlarged to the extent that the total system load has increased.
- SPECjbb2000 allows only one active Java Virtual Machine instance (JVM), whereas SPECjbb2005 permits several instances, which in turn achieves greater closeness to reality, particularly with large systems.

On the software side SPECjbb2005 measures the implementations of the JVM, JIT (Just-In-Time) compiler, garbage collection, threads and some aspects of the operating system. As far as hardware is concerned, it measures the efficiency of the CPUs and caches, the memory subsystem and the scalability of shared memory systems (SMP). Disk and network I/O are irrelevant.

SPECjbb2005 emulates a 3-tier client/server system that is typical for modern business process applications with emphasis on the middle tier system:

- Clients generate the load, consisting of driver threads, which on the basis of the TPC-C benchmark generate OLTP accesses to a database without thinking times.
- The middle-tier system implements the business processes and the updating of the database.
- The database takes on the data management and is emulated by Java objects that are in the memory. Transaction logging is implemented on an XML basis.

The major advantage of this benchmark is that it includes all three tiers that run together on a single host. The performance of the middle tier is measured, thus avoiding large-scale hardware installations and making direct comparisons possible between SPECjbb2005 results of different systems. Client and database emulation are also written in Java.

SPECjbb2005 only needs the operating system as well as a Java Virtual Machine with J2SE 5.0 features.

The scaling unit is a warehouse with approx. 25 MB Java objects. Precisely one Java thread per warehouse executes the operations on these objects. The business operations are assumed by TPC-C:

- New Order Entry
- Payment
- Order Status Inquiry
- Delivery
- Stock Level Supervision
- Customer Report

However, these are the only features SPECjbb2005 and TPC-C have in common. The results of the two benchmarks are not comparable.

SPECjbb2005 has 2 performance metrics:

- bops (business operations per second) is the overall rate of all business operations performed per second.
- bops/JVM is the ratio of the first metrics and the number of active JVM instances.

In comparisons of various SPECjbb2005 results it is necessary to state both metrics.

The following rules, according to which a compliant benchmark run has to be performed, are the basis for these metrics:

A compliant benchmark run consists of a sequence of measuring points with an increasing number of warehouses (and thus of threads) with the number in each case being increased by one warehouse. The run is started at one warehouse up through 2*MaxWhm but not less than 8 warehouses. MaxWhm is the number of warehouses with the highest operation rate per second the benchmark expects. Per default the benchmark equates MaxWH with the number of CPUs visible by the operating system.
The metrics bops is the arithmetic average of all measured operation rates with between MaxWhm warehouses and 2*MaxWhm warehouses.

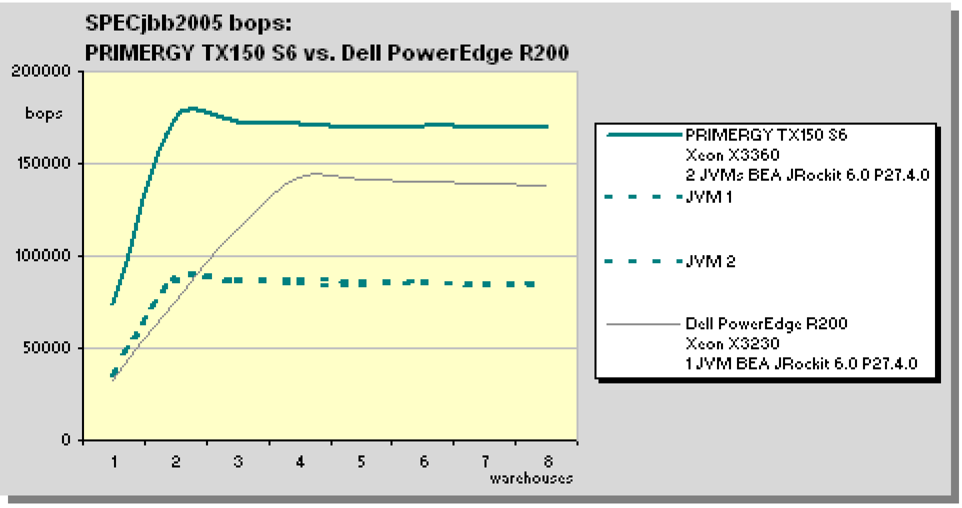
* SPECB®, SPECjbb® and the SPEC logo are registered trademarks of the Standard Performance Evaluation Corporation (SPEC).

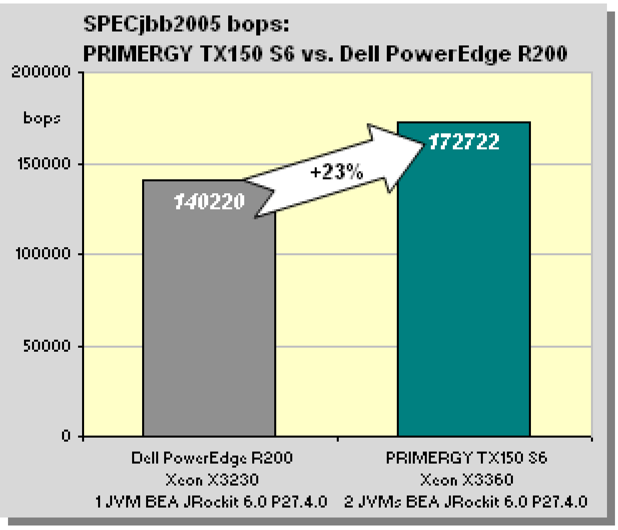
Benchmark results

In November 2007 the PRIMERGY TX150 S6 was measured with the Xeon X3210 processor and a memory of 8 GB PC2-6400 DDR2-SDRAM. The measurement was taken under Windows Server 2003 R2 Enterprise x64 Edition SP1. As JVM, two instances of JRockit(R) 6.0 P27.4.0 (build P27.4.0-10-90053-1.6.0_02-20071009-1827-windows-x86_64) by BEA were used.

In January 2008 the PRIMERGY TX150 S6 was measured in an otherwise unchanged configuration with the Xeon X3360 processor.

With the Xeon X3360 the PRIMERGY TX150 S6 achieved the best result of all mono-processor servers with a Quad-Core processor and exceeded the previous front runner in this category by 23%.* With the measurement of the PRIMERGY TX150 S6 all the measured values between 2 and 4 warehouses were incorporated in the overall benchmark result. With the measurement of the PowerEdge R200 this applies to all measured values between 4 and 8 warehouses.

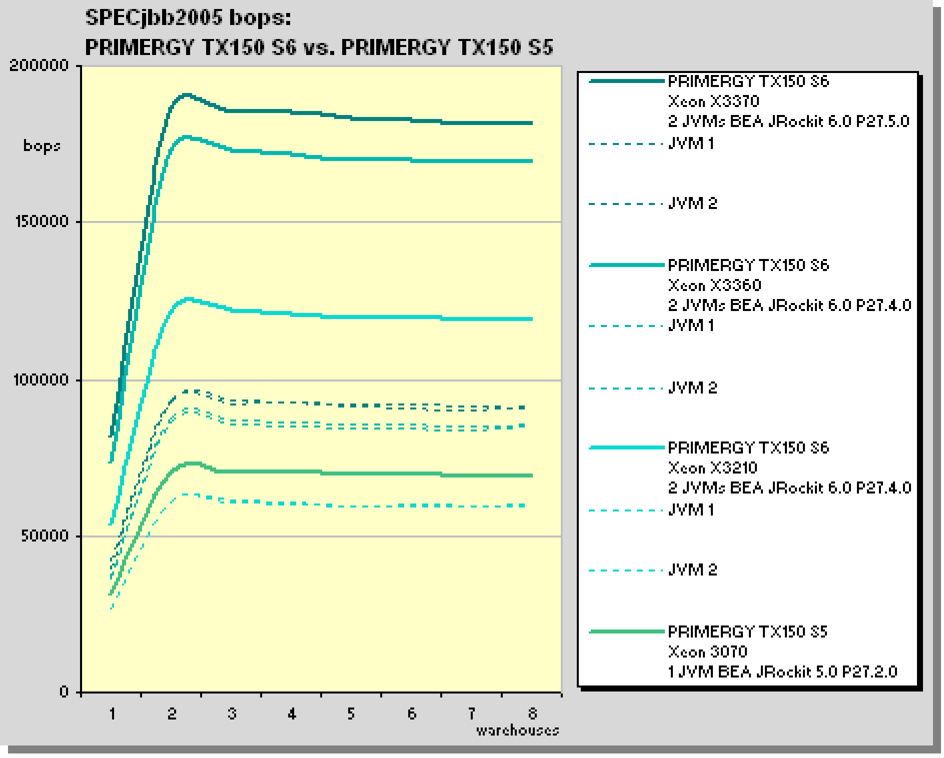




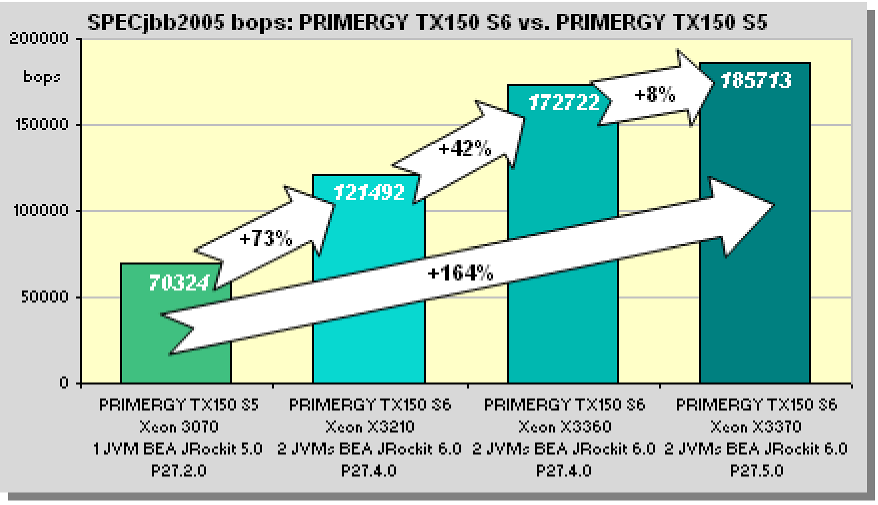
Source: http://www.spec.org/jbb2005/results, as of February 14, 2008

* Competitive benchmark results stated above reflect results published as of Feb 14, 2008. The comparison presented above is based on the best performing servers with one Quad-Core processor currently shipping by Dell and Fujitsu Siemens Computers, now operating under the name of Fujitsu. For the latest SPECjbb2005 benchmark results, visit http://www.spec.org/jbb2005/results.

In August 2008 the PRIMERGY TX150 S6 was measured with the Xeon X3370 processor and a memory of 8 GB PC2-6400 DDR2-SDRAM. The measurement was taken under Windows Server 2003 R2 Enterprise x64 Edition. As JVM, two instances of JRockit(R) 6.0 P27.5.0 (build P27.5.0-5-97156-1.6.0_03-20080403-1524-windows-x86_64) by BEA were used.



When comparing the PRIMERGY TX150 S6 and its predecessor the PRIMERGY TX150 S5 both in their highest performance configurations, an increase of +164% is noted. In all measurements the overall benchmark result includes the measured values from 2 to 4 warehouses.



Benchmark environment

The SPECjbb2005 measurements were performed on a PRIMERGY TX150 S6 with the following hardware and software configuration:

Hardware	
Model	PRIMERGY TX150 S6
CPU	Xeon X3210, X3360 and X3370
Number of chips	1 chip, 4 cores, 4 cores per chip
Primary Cache	32 kB instruction + 32 kB data on chip, per core
Secondary Cache	Xeon X3210: 8 MB (I+D) on chip, per chip Xeon X3360 and X3370: 12 MB (I+D) on chip, per chip
Other Cache	none
Memory	4 x 2 GB PC2-6400 DDR2-SDRAM
Software	
Operating System	Windows Server 2003 R2 Enterprise x64 Edition
JVM Version	Xeon X3210 and X3360: BEA JRockit(R) 6.0 P27.4.0 (build P27.4.0-10-90053-1.6.0_02-20071009-1827-windows-x86_64) Xeon X3370: BEA JRockit(R) 6.0 P27.5.0 (build P27.5.0-5-97156-1.6.0_03-20080403-1524-windows-x86_64)

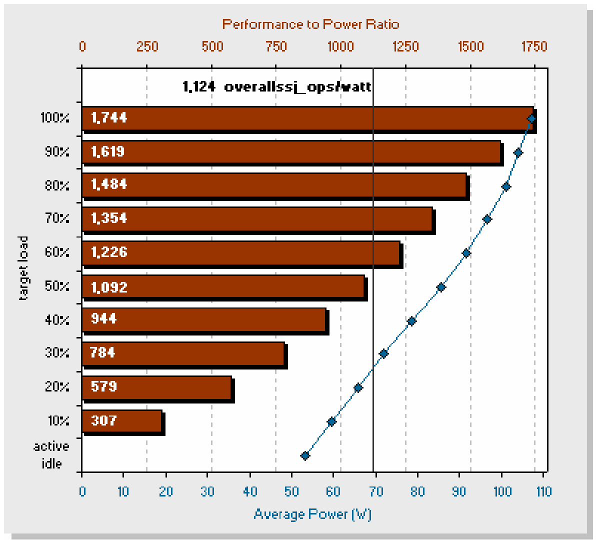
SPECpower_ssj2008*

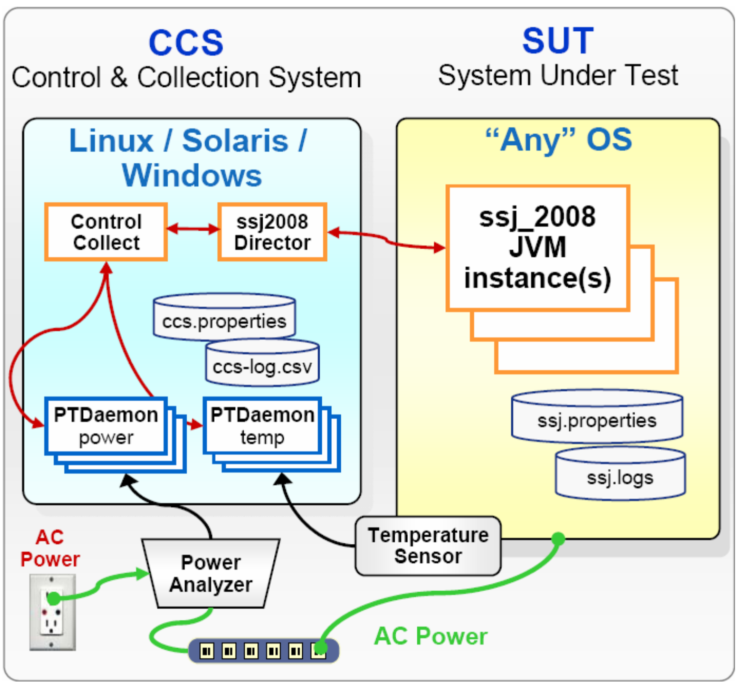
Benchmark description

SPECpower_ssj2008 is the first industry-standard SPEC benchmark that evaluates the power and performance characteristics of server class computers. With SPECpower_ssj2008, SPEC has defined server power measurement standards in the same way they have done for performance.

The benchmark workload represents typical server-side Java business applications. The workload is scalable, multi-threaded, portable across a wide range of operating environments, and economical to run. It exercises CPUs, caches, memory hierarchy, and the scalability of symmetric multiprocessor systems (SMPs), as well as implementations of the Java Virtual Machine (JVM), Just In Time (JIT) compiler, garbage collection, threads, and some aspects of the operating system.

SPECpower_ssj2008 reports power consumption for servers at different performance levels — from 100-percent to active idle in 10-percent segments — over a set period of time. The graduated workload recognizes the fact that processing loads and power consumption on servers vary substantially over the course of days or weeks. To compute a power-performance metric across all levels, measured transaction throughputs for each segment are added together, and then divided by the sum of the average power consumed for each segment. The result is a figure of merit called "overall ssj_ops/watt." This ratio gives information about the energy efficiency of the measured server. Because of its defined measurement standard it allows the customers to compare it to other configurations and servers measured with SPECpower_ssj2008. The adjoining diagram shows a typical graph of a SPECpower_ssj2008 result.





The benchmark runs on a wide variety of operating systems and hardware architectures and does not require extensive client or storage infrastructure. The minimum equipment for SPEC-compliant testing is two networked computers, plus a power analyzer and a temperature sensor. One computer is the System Under Test (SUT) running any of the supported operating systems along with the JVM installed. The JVM provides the environment required to run the SPECpower_ssj2008 workload which is implemented in Java. The other computer is a Collect and Control System (CCS) which controls the operation of the benchmark and captures the power, performance and temperature readings for reporting. The adjoining diagram gives an overview of the different components of this framework.

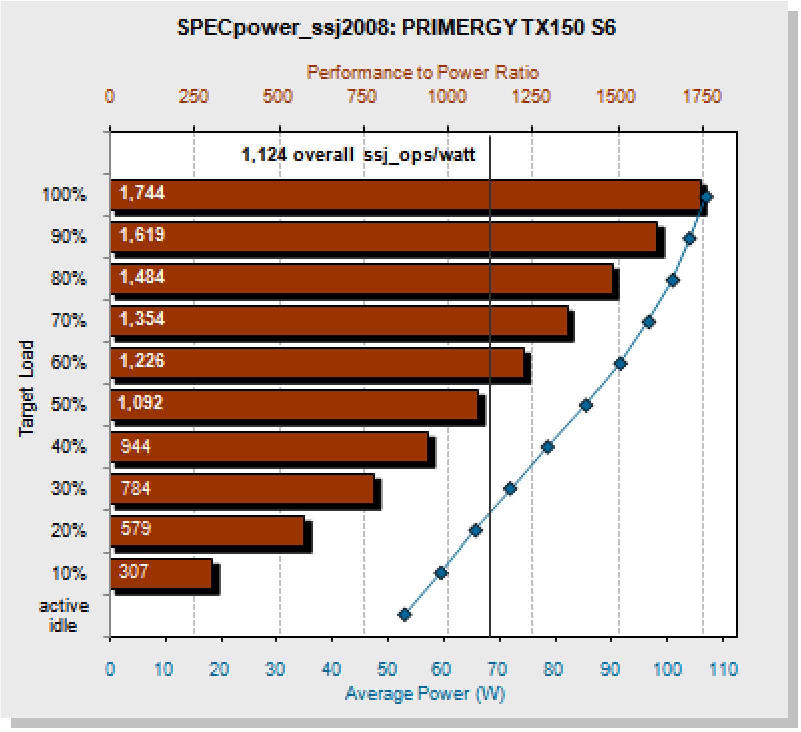
* SPEC®, SPECpower_ssj2008™ and the SPEC logo are registered trademarks of the Standard Performance Evaluation Corporation (SPEC).

Benchmark results

In June 2008 the PRIMERGY TX150 S6 was measured with an Intel Xeon X3360 processor and 4 GB of PC2-6400E DDR2-SDRAM memory. The measurement was taken under Windows Server 2003 R2 Enterprise x64 Edition and a JRockit(R) 6.0 P27.5.0 JVM by Oracle was used.

With the Xeon X3360 processor the PRIMERGY TX150 S6 achieved a world record score which exceeded the previous front runner by 6.7%* in energy efficiency. Compared to the measurement on the IBM System x3200 M2 which was performed with the same processor and achieved nearly the same throughput in ssj_ops this advance of the SPECpower_ssj2008 result of the PRIMERGY TX150 S6 is explained only by the lower power consumption at all load levels.

Additional tests with different configurations have been performed to show their influence on the server efficiency. Fujitsu does not submit all SPECpower_ssj2008 measurements for publication at SPEC. So, not all the results presented here appear on SPEC's web sites. But because we archive the results and log data for all measurements, we are able to prove the correct execution of the measurements at any time.



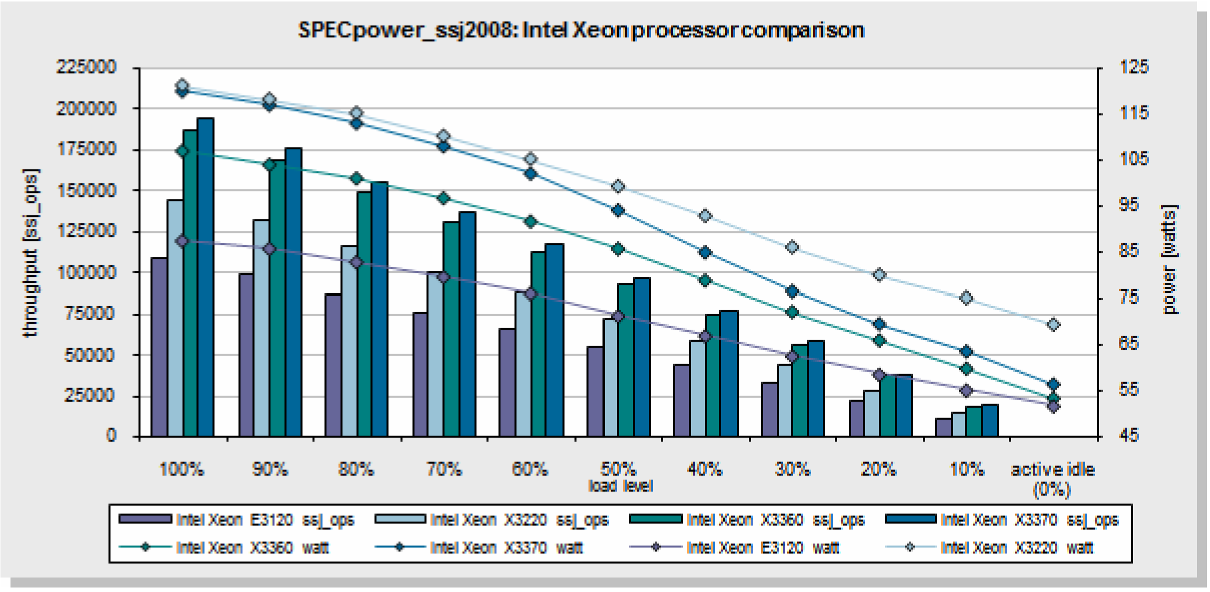
The adjoining diagram shows the result graph of the configuration described above, measured with the PRIMERGY TX150 S6. The red horizontal bars show the performance to power ratio in ssj_ops/watt (upper x-axis) for each target load level which are tagged on the y-axis of the diagram. The blue line shows the run of the curve for the average power consumption (bottom x-axis) at each target load level marked with a small rhomb.

The diagram shows how the efficiency of the server decreases with each target load level from 100% to active idle in 10% segments. The black vertical line shows the benchmark result of 1,124 overall ssj_ops/watt for the PRIMERGY TX150 S6. This is calculated by adding the measured transaction throughputs for each segment and then dividing by the sum of the average power consumed for each segment.

The configuration was tuned to get the best possible result for this server in terms of performance per watt. The memory configuration with 2 x 2 GB was selected to meet the criteria of best performance at lowest power consumption by populating only one slot of each available memory channel. This configuration enables the benchmark to use the full capacity of the available memory bandwidth and at the same time consume less power than a comparable performance equivalent 4 DIMMs configuration. We used the SATA base unit which has an integrated SATA controller in the chipset along with a 160 GB 3.5" SATA hard disk. Due to the absence of a dedicated onboard SAS controller which is included in the SAS base unit and the lower rotational speed of 7,200 rpm of the SATA hard disk this configuration was the best choice for this benchmark, because it consumes minimal power without impacting the performance. The most important factor in the hardware configuration is the right choice of the processor. Processors are the part of a server that consumes the most power beside the memory subsystem. For the PRIMERGY TX150 S6 the quad-core Xeon X3360 processor with a Thermal Design Power (TDP) of 95 watts showed the most efficient result. The influence of the different CPUs and other configuration options on the SPECpower_ssj2008 result is presented in subsequent diagrams.

* Competitive benchmark results stated above reflect results published as of Jun 4th, 2008. The comparison presented above is based on the most efficient servers currently shipping by IBM and Fujitsu Siemens Computers, now operating under the name of Fujitsu. For the latest SPECpower_ssj2008 benchmark results, visit http://www.spec.org/power_ssj2008/results.

The Xeon processor comparison diagram shows measurements with four different Xeon CPUs measured with the PRIMERGY TX150 S6. All other configuration details remained unchanged during the measurements.

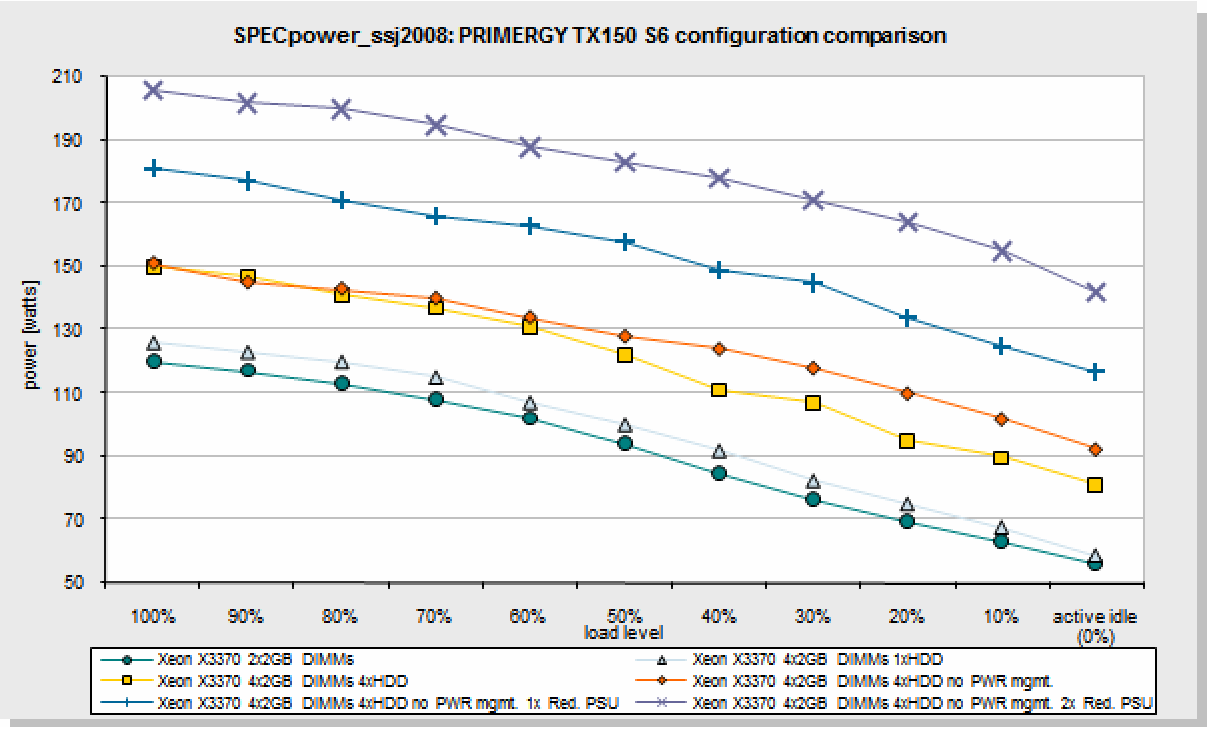


Obviously the throughput changes when using different CPUs with different frequencies or different number of cores per chip. That is exactly what the bars show in the diagram above. The measurement with the most powerful Xeon X3370 (quad-core 3.0 GHz) CPU delivers the highest throughput and the measurement with the Xeon E3120 (dual-core 3.16 GHz) CPU which has nearly the same frequency but half the number of cores per chip delivers the lowest throughput in ssj_ops (left y-axis). It is also visible that the ratio of the performance gain and the performance loss respectively with the different CPUs is nearly the same at each target load level (x-axis). But looking at the average power consumption curves (right y-axis) the behavior varies for the different load levels.

The ratio of power consumption between the different CPUs changes with every additional 10% of the target load. During active idle the difference is very small. That relates to the power management features of the CPUs and the operating system. They enable the CPUs to scale down the frequency and core voltage to a level where the CPUs consume the lowest power provided that the CPUs are idle. So the power consumption for the different CPUs is almost the same at active idle. As you can see this is not true for the Xeon X3220 processor. This is due to the fact that this processor is based on another manufacturing technology, 65 nm compared to 45 nm on all other processors. The 45 nm manufacturing technology enables the processors to consume less power and brings additional power management features. At the higher load levels the influence of the power management features is only marginal. This is exactly where the Xeon X3360 processor can play to its strength. Although it has the same TDP of 95 watts as the Xeon X3220 and X3370 the Xeon X3360 processor consumes much lower power at higher load levels. This is related to the lower frequency compared to the Xeon X3370 processor and to the less power consuming manufacturing technology of 45 nm compared to 65 nm in the case of the Xeon X3230 processor. At 100% load the power consumption difference to the Xeon X3370 processor is 23 watts.

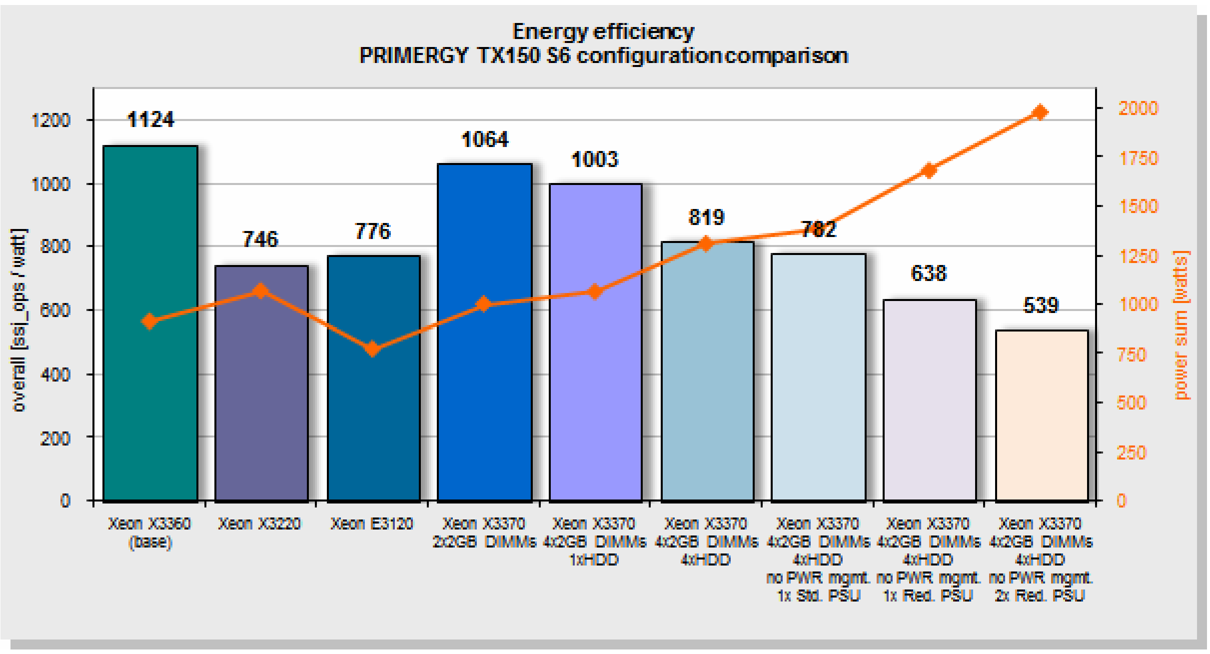
When looking at the Xeon E3120 processor you can see that it consumes the lowest power of all processors. But the lower power consumption of this CPU can not compensate the enormous performance drawback which is caused by the fact that this CPU has only 2 cores per chip. The Xeon X3360 processor shows the ideal balance between performance and power consumption and thus makes it the best choice for SPECpower_ssj2008.

The following diagram displays a comparison of additional configuration options. It only shows the differences in power consumption for each load level but no performance changes. The performance depends on the choice of the CPU as you could see in the diagram before. All the additional configuration changes we made in the following comparison, like adding more memory, more hard disks, other and additional Power Supply Units (PSUs) and turning off power management features, do not have any or only very limited influence on performance.



As you can see in the diagram above not only the right choice of processors has a significant influence on power consumption but other configuration details as well. Doubling the memory for example from 2 x 2 GB to 4 x 2 GB increases the power consumption at all load levels by about 6-7 watts on an average. Adding three additional hard disks to the configuration increases the power consumption by about 20-24 watts per load level on an average. When running the benchmark with all available power management features disabled in the BIOS and in the OS the PRIMERGY TX150 S6 consumes up to 15 watts more in the range of active idle to 50% load. The PRIMERGY TX150 S6 is available with two different types of PSUs. The first is a single standard PSU and the second are redundant PSUs. By replacing the standard PSU with one single redundant PSU the power consumption increases by up to 30 watts. This is explained by the fact that redundant power supplies need additional electronics (e.g. power backplane) for handling the redundancy and hot-plug functionality and the other important fact is that the redundant PSUs have a different efficiency compared to standard PSUs. When enabling full redundancy by adding the second redundant PSU the PRIMERGY TX150 S6 consumes another 25-30 watts more. The reason is that the PSUs share the load, i.e. each of them gets only half the load, and thus runs in a lower load range at a lower efficiency.

The final energy efficiency diagram shows the performance to power ratio (power efficiency) of all the previously mentioned configurations. It gives an overview of the benchmark result in overall ssj_ops/watt (left y-axis) each configuration achieved and how much power in watts (right y-axis) was consumed while running the benchmark.



As you can see the most energy efficient configuration does not consume the lowest power. It is still about 108% more efficient compared to the least energy efficient configuration. The configuration with the Xeon E3120 processor has the lowest power consumption. Although the best configuration with the Xeon X3360 processor does not deliver the highest throughput in ssj_ops and does not have the lowest power consumption, it achieves the highest benchmark score of 1,124 overall ssj_ops/watt.

The other results are important too, as they show the dependencies between the configurations and the efficiency. This information should give some hints about the power consumption and efficiency that can be expected from real world configurations used in customer installations.

Benchmark environment

All SPECpower_ssj2008 measurements presented here were performed on a PRIMERGY TX150 S6 with the following hardware and software configuration using the ZES Zimmer LMG95 power analyzer:

Hardware	
Model	PRIMERGY TX150 S6
Processor (TDP)	Xeon E3120 (65 W), X3220 (95 W), X3360 (95 W), X3370 (95 W)
Number of chips	1 chip, 2 cores per chip and 4 cores per chip
Primary Cache	32 KB instruction + 32 KB data on chip, per core
Secondary Cache	8 MB (I+D) on chip, per chip and 12 MB (I+D) on chip, per chip (6 MB shared / 2 cores)
Other Cache	none
Memory	2 x 2 GB PC2-6400E DDR2-SDRAM 4 x 2 GB PC2-6400E DDR2-SDRAM
Network Interface	1 x 1 GBit LAN Broadcom (onboard)
Disk Subsystem	1 × Integrated SATA Controller 1 × 3.5” SATA disk, 160 GB, 7.2 krpm, JBOD 4 × 3.5” SATA disk, 160 GB, 7.2 krpm, JBOD
Power Supply Unit	1 x 350 W DPS-350UB A 1 x 450 W DPS-450FB G 2 x 450 W DPS-450FB G
Software	
Operating System	Windows Server 2003 R2 Enterprise x64 Edition
JVM Version	Oracle JRockit(R) 6.0 P27.5.0 (build P27.5.0-5_o_CR371811_CR374296-100684-1.6.0_03-20080702-1651-windows-x86_64)
JVM options	-Xms1650m -Xmx1650m -Xns1500m -XXaggressive -Xlargepages -Xgc:genpar -XXcallprofiling -XXgcthreads=2 -XXtlasize:min=4k,preferred=1024k -XXthroughputcompaction

SPECweb2005*

Benchmark description

SPECweb2005 is the next generation web server benchmark developed by the Open Systems Group (OSG) of the Standard Performance Evaluation Corporation (SPEC). It is the successor of SPECweb99 and SPECweb99_SSL and it measures the performance of a HTTP server under a standardized load of static and dynamic requests. The new version includes many sophisticated and state-of-the-art enhancements to meet the modern demands of Web users of today and tomorrow.

Contrary to its predecessor version, SPECweb2005 is split into three different workloads, which are based on real-world web-server applications:

- **SPECweb2005_Banking** – Emulates typical online banking requests, such as login/logoff, account status, bank transfers, displaying and changing user profiles, etc. Login includes the setting up an SSL connection that will be used for all following activities.
- **SPECweb2005_Ecommerce** – Simulates an online transaction in the computer business. Users can look through the pages, view goods, put them in their shopping carts and purchase the products. Activities in the initial phases of the connection use non-encrypted connections. As soon as an order is to be sent off, the connections are SSL-encrypted.
- **SPECweb2005_Support** – Emulates requests coming in on a support web site. Users can search through the page, view lists of available products and download the related files. Requests are always non-encrypted.

The requests of all three workloads refer to dynamically generated contents and static files of various sizes. Intervals between requests ("think times") vary. The distribution of the requests and the think times are controlled by tables and functions. Average values for these parameters are laid down in configuration files and are monitored by the sequencing unit.

SPECweb2005 is not tied to a particular operating system or to a particular web server. The benchmark environment consists of several components. Each client system runs a load generator program setting up connections to the web server, sending page requests and receiving web pages in response to the requests. A prime client initializes the other systems, monitors the test procedure, collects the results and evaluates them. The web server, also referred to as "System Under Test" (SUT), comprises the hardware and software used to handle the requests. A new feature is the back-end simulator (BeSim) that emulates the database and application components of the entire application. The web server communicates with the BeSim via HTTP requests to obtain any additional information required. The sequencer and the client programs are written in Java and are divided into individual threads, each of which emulates a virtual user session.

All three workloads pass various phases during the test. In the ramp-up phase, the load-generating threads are started one after another. This is followed by a warm-up phase initializing the measurement. Any previously recorded results and errors are deleted before the actual measuring interval begins. During the measuring phase all requests and responses are recorded in the final results. In the ramp-down phase which now follows the threads are stopped, followed by an idle phase, before the next test iteration begins with another ramp-up phase. Thus altogether three iterations are performed for each workload.

The number of generated threads is defined separately for each workload, according to the performance of the SUT in the test configuration. To determine the results, the clients measure for each requested page the time between the sending of the request and the arrival of all the data of the requested page. The response times for embedded image files are also included in the calculation. The result takes all those pages into account that meet particular QoS (Quality of Service) criteria. For this purpose the responses are assigned to the following categories according to response times (Banking and Ecommerce) and transfer rates (Support) within the workloads:

- GOOD – response time < 2s (Banking), < 3s (Ecommerce); transfer rate > 99000 bytes/s (Support)
- TOLERABLE – response time < 4s (Banking), < 5s (Ecommerce); transfer rate > 95000 bytes/s (Support)
- FAILED – response time > 4s (Banking), > 5s (Ecommerce); transfer rate < 95000 bytes/s (Support)

In all three test iterations at least 95% of all responses must fall into category GOOD and 99% into category TOLERABLE for the workload result to be valid. A regular overall result requires valid partial results in all three workloads with the same system configuration.

The individual results are named after the workloads and indicate the maximum number of user sessions that can be handled by the system under test with the QoS criteria being met. They thus allow a system to be assessed under different realistic conditions. To calculate the overall result, each partial result is related to a reference value; then the geometric mean of these three values is calculated, multiplied by 100. The overall result (**SPECweb2005**) thus indicates the relative performance of the measured system in relation to the reference system.

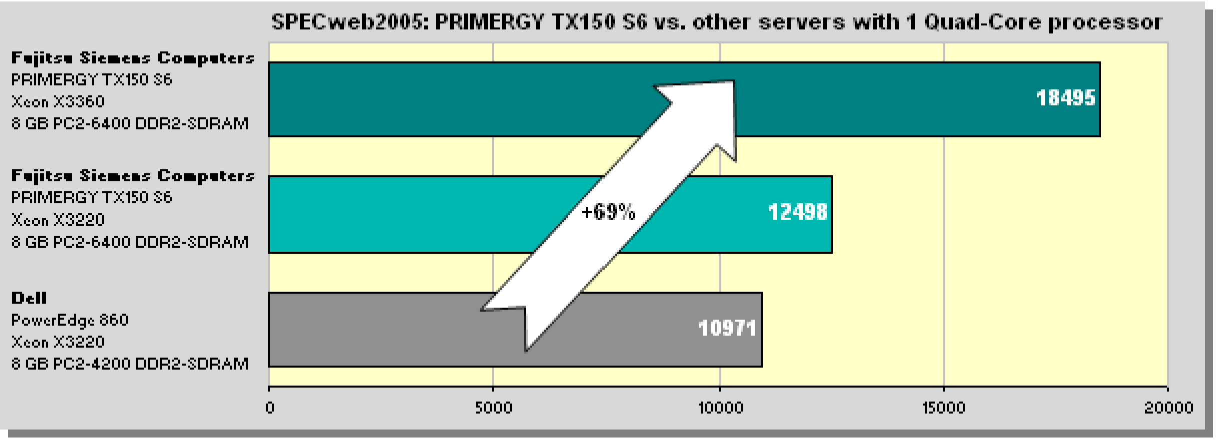
* SPEC®, SPECweb® and the SPEC logo are registered trademarks of the Standard Performance Evaluation Corporation (SPEC).

Benchmark results

In April 2008 the PRIMERGY TX150 S6 was measured with one Xeon X3220 processor and 8 GB PC2-6400 DDR2-SDRAM. Two quad-port Intel PRO/1000GT and one Broadcom NetXtreme II BCM5708 (onboard) were used for the network. A FibreCAT CX500 with 45 hard disks, which was connected via an Emulex LP10000DC fibre channel controller, was used as disk subsystem. A RAID 5 was built across the 45 hard disks. The logging was done on a hard disk of type Seagate ST380013AS. The operating system was resident on a Seagate ST3160815AS hard disk. Both hard disks were connected with the onboard SATA controller. The measurement was performed using the HTTP software Accoria Rock Web Server v1.4.6 (x86_64) under Red Hat Enterprise Linux 5.1 (2.6.18-53.el5 x86_64).

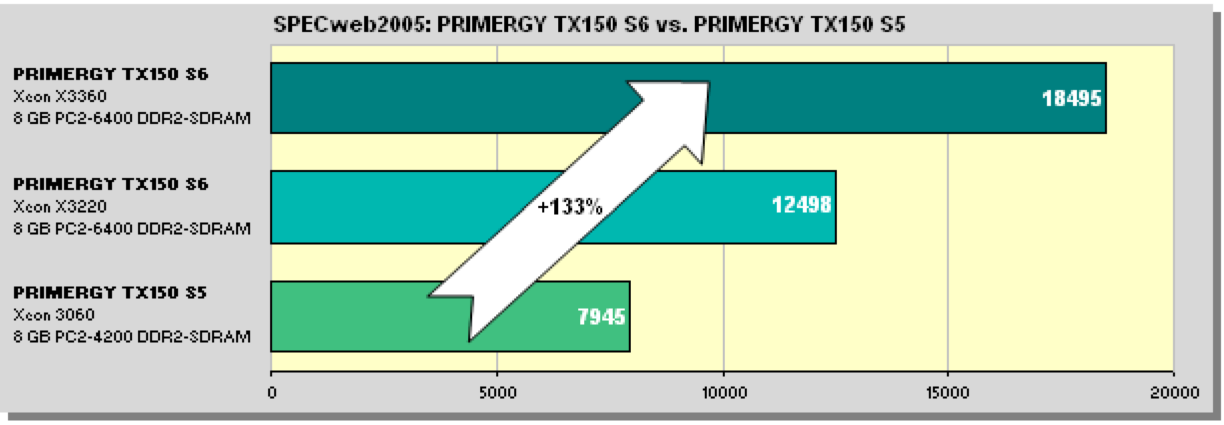
In May 2008 the PRIMERGY TX150 S6 was measured with one Xeon X3360 processor and 8 GB PC2-6400 DDR2-SDRAM. Two quad-port Intel PRO/1000GT and one Broadcom NetXtreme II BCM5708 (onboard) were used for the network. Two FibreCAT SX88 with 24 hard disks, which were connected via an Emulex LPe11002 fibre channel controller, were used as disk subsystem. A RAID 0 was built across the 24 hard disks. The logging was done on a hard disk of type Seagate ST380013AS. The operating system was resident on a Seagate ST3160815AS hard disk. Both hard disks were connected with the onboard SATA controller. The measurement was performed using the HTTP software Accoria Rock Web Server v1.4.6 (x86_64) under Red Hat Enterprise Linux 5.1 (2.6.18-53.el5 x86_64).

In the class of servers with one Quad-Core processor, the PRIMERGY TX150 S6 achieved the best result.*



Source: http://www.spec.org/web2005/results/, as of August 5, 2008

Compared with the PRIMERGY TX150 S5 the PRIMERGY TX150 S6 improved the throughput performance by 133%.



* Competitive benchmark results stated above reflect results published as of August 5, 2008. The comparison presented above is based on the best performing servers with one Quad-Core processor currently shipping by Dell and Fujitsu Siemens Computers, now operating under the name of Fujitsu. For the latest SPECweb2005 benchmark results, visit http://www.spec.org/web2005/results/.

Benchmark environment

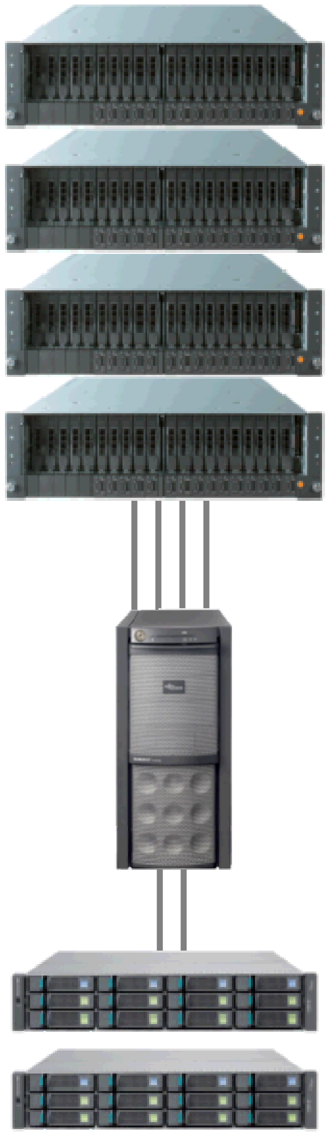
TX150 S6 with Xeon X3220:

60 × PRIMERGY BX300
2 x Pentium III 933 MHz
1 GB RAM
2 x Broadcom NetXtreme (onboard)
Windows XP Professional SP1

PRIMERGY TX150 S6
1 × Xeon 3220
8 GB PC2-6400 DDR2-SDRAM
1 × Emulex LP10000DC fibre channel controller
2 × dual-channel Intel PRO/1000GT
1 × Broadcom NetXtreme II BCM 5708 (onboard)
Operating system: Red Hat Enterprise Linux 5.1
(2.6.18-53.el5 x86_64)
HTTP software: Accoria Rock Web Server v1.4.6 (x86_64)

Disk subsystem
1 × FibreCAT CX500 with
45 disks

TX150 S6 with Xeon X3360:



80 × PRIMERGY BX300
2 x Pentium III 933 MHz
1 GB RAM
2 x Broadcom NetXtreme (onboard)
Windows XP Professional SP1

PRIMERGY TX150 S6
1 × Xeon 3360
8 GB PC2-6400 DDR2-SDRAM
1 × Emulex LPe11002 fibre channel controller
2 × dual-channel Intel PRO/1000GT
1 × Broadcom NetXtreme II BCM 5708 (onboard)
Operating system: Red Hat Enterprise Linux 5.1 (2.6.18-53.el5 x86_64)
HTTP software: Accoria Rock Web Server v1.4.6 (x86_64)

Disk subsystem
2 × FibreCAT SX88 with
24 disks

StorageBench

Benchmark description

To estimate the capability of disk subsystems Fujitsu Technology Solutions defined a benchmark called StorageBench to compare the different storage systems connected to a system. To do this StorageBench makes use of the Iometer measuring tool developed by Intel combined with a defined set of load profiles that occur in real customer applications and a defined measuring scenario.

Measuring tool

Since the end of 2001 Iometer has been a project at http://SourceForge.net and is ported to various platforms and enhanced by a group of international developers. Iometer consists of a user interface for Windows systems and the so-called "dynamo" which is available for various platforms. For some years now it has been possible to download these two components under "Intel Open Source License" from http://www.iometer.org/ or http://sourceforge.net/projects/iometer.

Iometer gives you the opportunity to reproduce the behavior of real applications as far as accesses to IO subsystems are concerned. For this purpose, you can among other things configure the block sizes to be used, the type of access, such as sequential read or write, random read or write and also combinations of these. As a result Iometer provides a text file with comma separated values (.csv) containing basic parameters, such as throughput per second, transactions per second and average response time for the respective access pattern. This method permits the efficiency of various subsystems with certain access patterns to be compared. Iometer is in a position to access not only subsystems with a file system, but also so-called raw devices.

With Iometer it is possible to simulate and measure the access patterns of various applications, but the file cache of the operating system remains disregarded and operation is in blocks on a single test file.

Load profile

The manner in which applications access the mass storage system considerably influences the performance of a storage system. Examples of various access patterns of a number of applications:

Application	Access pattern
Database (data transfer)	random, 67% read, 33% write, 8 KB (SQL Server)
Database (log file)	sequential, 100% write, 64 KB blocks
Backup	sequential, 100% read, 64 KB blocks
Restore	sequential, 100% write, 64 KB blocks
Video streaming	sequential, 100% read, blocks ≥ 64 KB
File server	random, 67% read, 33% write, 64 KB blocks
Web server	random, 100% read, 64 KB blocks
Operating system	random, 40% read, 60% write, blocks ≥ 4 KB
File copy	random, 50% read, 50% write, 64 KB blocks

From this four distinctive profiles were derived:

Load profile	Access	Access pattern		Block size	Load tool
		read	write		
Streaming	sequential	100%		64 KB	Iometer
Restore	sequential		100%	64 KB	Iometer
Database	random	67%	33%	8 KB	Iometer
File server	random	67%	33%	64 KB	Iometer

All four profiles were generated with Iometer.

Measurement scenario

In order to obtain comparable measurement results it is important to perform all the measurements in identical, reproducible environments. This is why StorageBench is based, in addition to the load profile described above, on the following regulations:

- Since real-life customer configurations work only in exceptional situations with raw devices, performance measurements of internal disks are always conducted on disks containing file systems. NTFS is used for Windows and ext3 for Linux, even if higher performance could possibly be achieved with other file systems or raw devices.
- Hard disks are among the most error-prone components of a computer system. This is why RAID controllers are used in server systems in order to prevent data loss through hard disk failure. Here several hard disks are put together to form a "Redundant Array of Independent Disks", known as RAID in short – with the data being spread over several hard disks in such a way that all the data is retained even if one hard disk fails – except with RAID 0. The most usual methods of organizing hard disks in arrays are the RAID levels RAID 0, RAID 1, RAID 5, RAID 6, RAID 10, RAID 50 and RAID 60. Information about the basics of various RAID arrays is to be found in the paper Performance Report - Modular RAID for PRIMERGY.

  Depending on the number of disks and the installed controller, the possible RAID configurations are used for the StorageBench analyses of the PRIMERGY servers. For systems with two hard disks we use RAID 1 and RAID 0, for three and more hard disks we also use RAID 1E and RAID 5 and, where applicable, further RAID levels – provided that the controller supports these RAID levels.
- Regardless of the size of the hard disk, a measurement file with the size of 8 GB is always used for the measurement.
- In the evaluation of the efficiency of I/O subsystems, processor performance and memory configuration do not play a significant role in today's systems - a possible bottleneck usually affects the hard disks and the RAID controller, and not CPU and memory. Therefore, various configuration alternatives with CPU and memory need not be analyzed under StorageBench.

Measurement results

For each load profile StorageBench provides various key indicators: e.g. "data throughput" in megabytes per second, in short MB/s, "transaction rate" in I/O operations per second, in short IO/s, and "latency time" or also "mean access time" in ms. For sequential load profiles data throughput is the normal indicator, whereas for random load profiles with their small block sizes the transaction rate is normally used. Throughput and transaction rate are directly proportional to each other and can be calculated according to the formula

Data throughput [MB/s]	*= Transaction rate [Disk-I/O s^{-1}] × Block size [MB]*
Transaction rate [Disk-I/O s^{-1}]	*= Data throughput [MB/s] / Block size [MB]*

Benchmark results

The PRIMERGY TX150 S6 is equipped with controllers from the “Modular RAID” family. The variety of the RAID solutions enables the user to choose the right controller for his application scenario.

The PRIMERGY TX150 S6 has the following RAID solutions to offer:

1. SATA RAID onboard controller

   The controller is implemented directly on the motherboard of the server in the Intel ICH9R chipset and the RAID stack is realized by the server CPU. This RAID solution is only foreseen for the connection of SATA hard disks. Support is provided for RAID levels 0, 1 and 10 as well as for RAID 5 with an additional iButton. This controller does not have a controller cache.

2. RAID Controller LSI MegaRAID SAS 1068

   The controller is supplied as a PCI Express card. The maximum number of SATA and SAS hard disks that can be connected to the controller is eight. Support is provided for RAID levels 0, 1 and 1E. The controller does not have a cache.

3. RAID Controller LSI MegaRAID SAS 1078

   The controller is supplied as a PCI Express card and offers the user a complete RAID solution. Both SATA and SAS hard disks can be connected. Support is provided for RAID levels 0, 1, 5, 6, 10, 50 and 60. Two different versions of this controller are on offer with either a 256 MB or 512 MB cache. The controller cache can be protected against power failure by an optional battery backup unit (BBU). The controller supports up to 240 hard disks.

Various SATA and SAS hard disks can be connected to these controllers. Depending on the performance required, it is possible to select the appropriate disk subsystem. Depending on the model version the PRIMERGY TX150 S6 offers four 3½" SAS/SATA or eight 2½" SAS hot-plug bays for hard disks. Optionally, an extension box is available for the 3½" variant with two additional 3½" hot-plug bays.

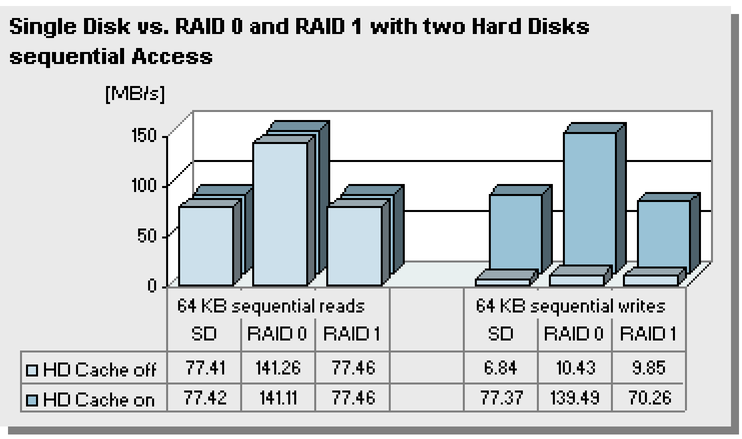
The following hard disks can be chosen for the PRIMERGY TX150 S6:

- 2½" SAS hard disks with a capacity of 36 GB, 73 GB and 146 GB (10 krpm)
- 2½" SAS hard disks with a capacity of 36 GB and 73 GB (15 krpm)
- 3½" SAS hard disks with a capacity of 73 GB, 146 GB and 300 GB (10 krpm)
- 3½" SAS hard disks with a capacity of 73 GB, 146 GB and 300 GB (15 krpm)
- 3½" SATA hard disks with a capacity of 160 GB, 250 GB, 500 GB and 750 GB (7.2 krpm)

The hard disk cache has influence on disk I/O performance. Unfortunately, this is frequently seen as a security problem in the event of a power failure and is therefore disabled. On the other hand, it was for a good reason integrated by the hard disk manufacturers to increase write performance. Features, such as Native Command Queuing (NCQ), only function at all when the disk cache is enabled. For performance reasons it is advisable to enable the disk cache for the SATA hard disks in particular, which in comparison with the SAS hard disks rotate slowly. The by far larger cache for I/O accesses and thus a potential security risk for data loss in the event of a power failure is in any case in the main memory and is administered by the operating system. To prevent data losses it is advisable to equip the system with an uninterruptible power supply (UPS).

SATA RAID Onboard Controller ICH9R

The following illustrations use 3½" SATA hard disks to show how throughput depends on cache settings. The throughputs of a single hard disk (Single Disk, SD) are compared with the throughputs of two hard disks in a RAID 0 and RAID 1 array.



Read throughput for sequential reading of 64 KB blocks is not dependent on the cache settings. In RAID 1 roughly the same throughput values are achieved as in the single disk configuration, although RAID 1 offers the benefit of data redundancy. RAID 0 has a better utilization of capacity and two hard disks RAID 0 almost double the read throughput.

In contrast, write throughput with sequential access with 64 KB blocks largely depends on the cache settings. Enabling the disk cache improves the write throughput by a factor of about 11 in a single disk configuration, by a factor of 13 in RAID 0 and by a factor of 7 in RAID 1. The much higher write throughput is explained by the optimized write accesses to the hard disk and

the shorter latency times. Here again the RAID 0 array achieves almost twice the write throughput through parallel accesses, as compared with the other two configurations.

Enabling the disk cache leads to an increase in throughput during random access. However this increase is not as noticeable as with sequential writing. With a random access with 64 KB blocks and a single disk configuration the increase in throughput is about 36%, in RAID 0 about 22% and in RAID 1 about 19%.

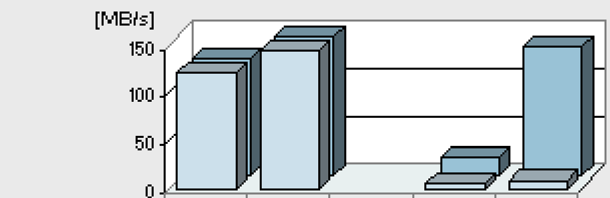
In the case of random access with 8 KB blocks the increase in throughput is a little higher than in the case with random access with 64 KB blocks and is roughly 23% with single disk, 36% with RAID 0 and about 20% with RAID 1.

Single Disk vs. RAID 0 and RAID 1 with two Hard Disks random Access

	8 KB random 67% reads				64 KB random 67% reads		
	SD	RAID 0	RAID 1		SD	RAID 0	RAID 1
HD Cache off	0.74	1.13	0.81		6.30	6.25	5.57
HD Cache on	0.91	1.54	0.97		8.57	7.60	6.63

The onboard SATA RAID ICH9R controller does not have a controller cache. This fact becomes particularly evident during write access in RAID 5. The enabling of this disk cache only brings about a moderate increase in throughput of about three-fold. Measurements with the LSI MegaRAID SAS 1078 controller, which has a controller cache, have shown that as a result of the enabled controller cache throughput increases of up to 39-fold and more are possible. On the other hand with RAID 10, which actually consists of two »striped« RAID 1 arrays, the enabling of the disk cache fully benefits write throughput, which increases by about 13-fold.

RAID 5 and RAID 10 with four Hard Disks sequential Access



	64 KB sequential reads			64 KB sequential writes	
	RAID 5	RAID 10		RAID 5	RAID 10
HD Cache off	124.49	147.12		6.65	10.40
HD Cache on	124.22	147.15		20.53	137.18

In contrast, the disk cache has no impact on the throughputs for sequential read with 64 KB blocks.

In RAID 10 better throughputs are achieved than in RAID 5. For sequential read access with 64 KB blocks and enabled disk cache the throughput is about 18% higher. The throughput is about 6.7-fold higher for sequential write access with 64 KB blocks and enabled disk cache. The disadvantage of RAID 10 compared with RAID 5 lies in the poorer capacity utilization. The loss of capacity in a configuration with four hard disks is 50% with RAID 10 and only 25% with RAID 5.

The throughput differences between RAID 5 and RAID 10 are also evident during random access. However, these differences are not as prominent as with sequential write. The throughput difference between RAID 5 and RAID 10 depends on the block size. For random access with 8 KB blocks the throughput in RAID 10 is 25% higher and with 64 KB blocks about 14% higher than in RAID 5. For random access the enabling of the disk cache brings about a throughput increase of between 27% and 39% with RAID 10 and about 24% with RAID 5. The slightly poorer throughputs measured in the RAID 5 array can be explained through the additional outlay required during the creation of the parity block.

RAID 5 and RAID 10 with four Hard Disks random Access

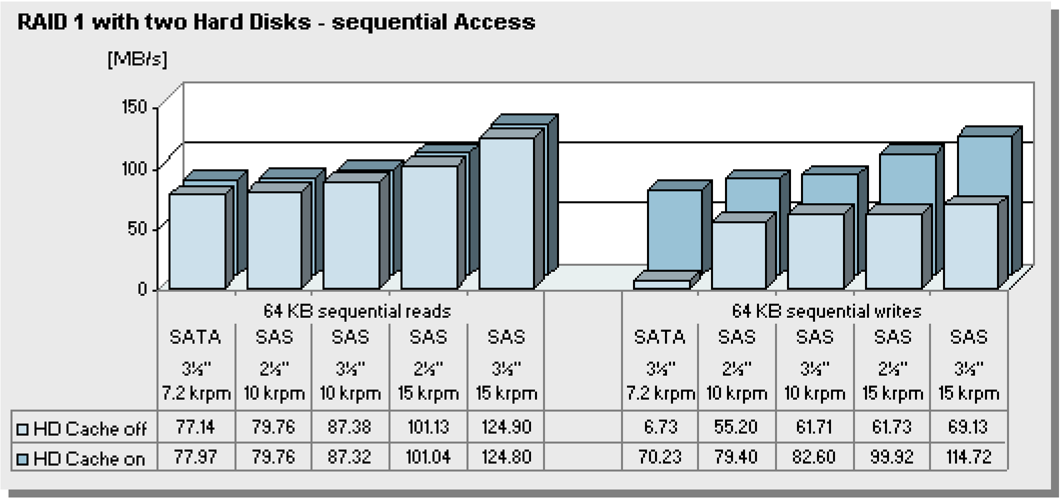
	8 KB random 67% reads			64 KB random 67% reads	
	RAID 5	RAID 10		RAID 5	RAID 10
HD Cache off	0.84	0.93		4.75	5.27
HD Cache on	1.03	1.30		5.89	6.69

LSI MegaRAID SAS 1068

Below is a comparison of the performance of various hard disk types with the LSI MegaRAID SAS 1068 controller. The controller does not have a controller cache. Therefore, measurements were only performed with and without a disk cache.

In the test setup two hard disks were connected to the controller and configured as a RAID 1. In the measurements all the hard disk types currently available to the PRIMERGY TX150 S6 were analyzed. The throughputs of the individual hard disks types in RAID 1 are compared below with different access patterns.

The diagram shows that as the rotational speed increases, the throughput for sequential reads and writes with a 64 KB block size rises.



If for sequential read with enabled disk cache a hard disk with a rotational speed of 15 krpm is used instead of one with a speed of 10 krpm, the result for the 2½" hard disk is an increase in throughput of about 27% and about 43% for the 3½" hard disk. If you compare the throughputs of the 2½" and 3½" hard disks both with a rotational speed of 10 krpm, you can see that the throughput for the 3½" hard disk is about 9% higher than for the 2½" hard disk. At a rotational speed of 15 krpm the difference in throughput between the 2½" and 3½" hard disk is even greater and amounts to 23%.

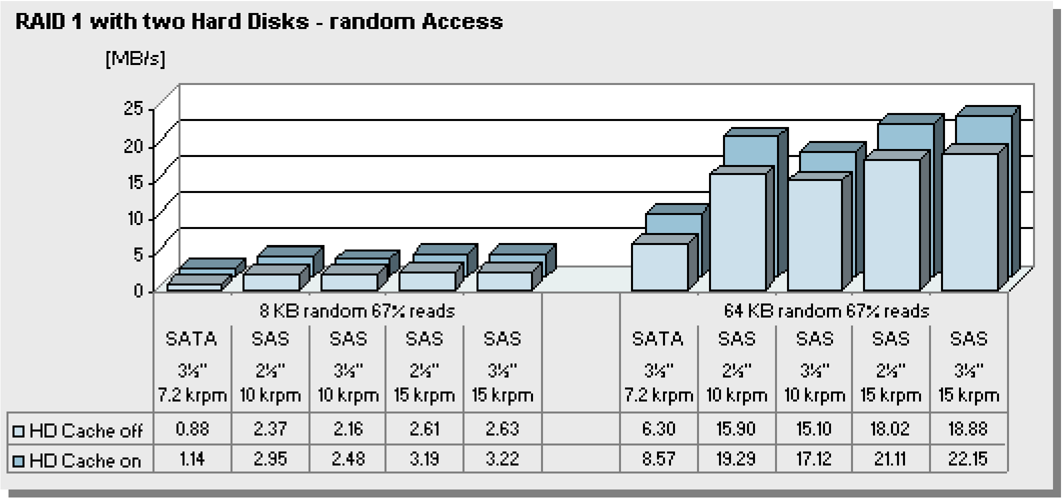
If you compare the 3½" SAS hard disk with the 3½" SATA hard disk, you can then see that the throughput of the SAS hard disk with 10 krpm is about 12% higher than the SATA hard disk with 7.2 krpm for sequential read and with an enabled disk cache. If you compare the 3½" SAS hard disk with 15 krpm with the SATA hard disk, you see that the throughput of the 3½" SAS hard disk with 15 krpm is even 60% higher than with the SATA hard disk.

If for sequential write with enabled disk cache a hard disk with a rotational speed of 15 krpm is used instead of one with a speed of 10 krpm, the result for the 2½" hard disk is an increase in throughput of about 26% and about 39% for the 3½" hard disk. If you compare the throughputs of the 2½" and 3½" hard disks both with a rotational speed of 10 krpm, you can see that the throughput for the 3½" hard disk is about 4% higher than for the 2½" hard disk. At a rotational speed of 15 krpm the difference in throughput between the 2½" and 3½" hard disk is even greater and amounts to 15%.

If you compare the 3½" SAS hard disk with the 3½" SATA hard disk, you can then see that the throughput of the SAS hard disk with 10 krpm is about 18% higher than the SATA hard disk with 7.2 krpm for sequential write and with an enabled disk cache. If you compare the 3½" SAS hard disk with 15 krpm with the same SATA hard disk, you see that the throughput of the 3½" SAS hard disk with 15 krpm is even 63% higher than with the SATA hard disk.

A special increase in throughput for sequential write, up to 10.4-fold, can be achieved with the SATA hard disk by enabling the disk cache. The increase in throughput gained with SAS hard disks through enabling the disk cache is not so pronounced as with the SATA hard disks, but it is still significant. For the 2½" hard disks with 10 krpm throughput increases by 44% and by about 62% for the 2½" hard disks with 15 krpm. For the 3½" hard disks with 10 krpm throughput increases by 34% and by about 66% for the 3½" hard disks with 15 krpm.

The following diagram shows that for random access with 67% reads the disk cache also plays an important role in improving throughput.



	SATA 3½" 7.2 krpm	SAS 2½" 10 krpm	SAS 3½" 10 krpm	SAS 2½" 15 krpm	SAS 3½" 15 krpm		SATA 3½" 7.2 krpm	SAS 2½" 10 krpm	SAS 3½" 10 krpm	SAS 2½" 15 krpm	SAS 3½" 15 krpm
HD Cache off	0.88	2.37	2.16	2.61	2.63		6.30	15.90	15.10	18.02	18.88
HD Cache on	1.14	2.95	2.48	3.19	3.22		8.57	19.29	17.12	21.11	22.15

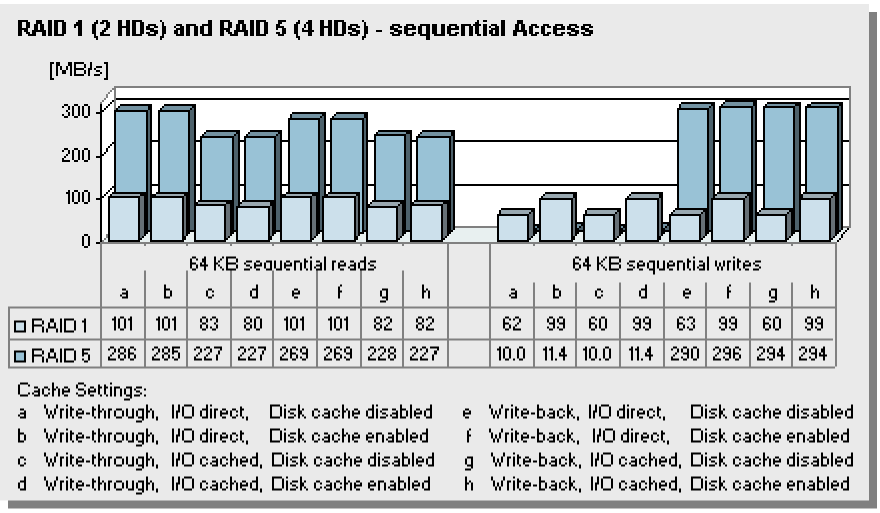
An increase in throughput of up to 30% was achieved for the SATA hard disk with 8 KB blocks. For the 3½" 15 krpm SAS hard disks the increase in throughput with 8 KB blocks is somewhat smaller, namely 23%. If you compare the throughput of the 3½" SAS hard disk with the 3½" SATA hard disk, you can then see that the throughput of the SAS hard disk with 10 krpm is about 2.2 times higher for random access with 8 KB blocks and enabled disk cache than in the SATA hard disk with 7.2 krpm. If you compare the throughput of the 3½" SAS hard disk with 15 krpm with the 3½" SATA hard disk with 7.2 krpm, you can then see that the throughput of the 3½" SAS hard disk is about 2.8 times higher for random access with 8 KB blocks and enabled disk cache than with the SATA hard disk.

LSI MegaRAID SAS 1078

The RAID array defines the way in which data is treated as regards availability. How quickly the data is transferred in the respective RAID array context depends largely on the data throughput of the hard disks. The number of hard disks configured for the measurements in a RAID array was defined depending on the RAID level. Between two and four hard disks were used. To ensure that the hard disks do not represent a bottleneck when determining the efficiency of the controller under various cache settings, the measurements were performed with 2½" hard disks with a rotational speed of 15 krpm.

Throughput can in part be considerably increased through the cache settings. However, these increases in throughput differ – depending on the data structure and access pattern. For the measurements the controller cache option "Read-Mode" is always set to "No Read-ahead". The options "Write-Mode", "I/O cache" and "Disk cache" were varied.

The following diagram shows the throughputs for sequential read and write with 64 KB blocks and for different cache settings in RAID 1 with two and in RAID 5 with four 2½" SAS hard disks. The read throughput in RAID 1 with optimum cache settings is in the range of the maximum possible throughput of over 100 MB/s. The "I/O cached" cache option has a negative impact on read throughput in RAID 1.



	a	b	c	d	e	f	g	h		a	b	c	d	e	f	g	h
RAID 1	101	101	83	80	101	101	82	82		62	99	60	99	63	99	60	99
RAID 5	286	285	227	227	269	269	228	227		10.0	11.4	10.0	11.4	290	296	294	294

Cache Settings:

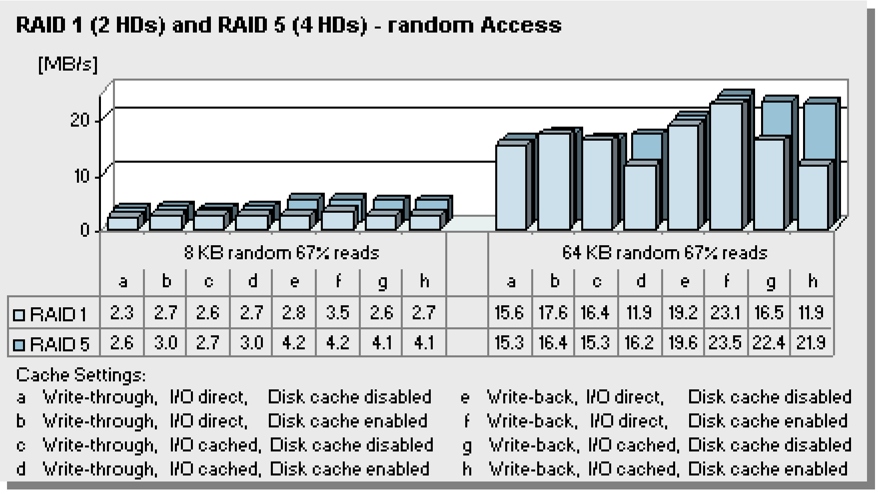
a	Write-through,	I/O direct,	Disk cache disabled	e	Write-back,	I/O direct,	Disk cache disabled
b	Write-through,	I/O direct,	Disk cache enabled	f	Write-back,	I/O direct,	Disk cache enabled
c	Write-through,	I/O cached,	Disk cache disabled	g	Write-back,	I/O cached,	Disk cache disabled
d	Write-through,	I/O cached,	Disk cache enabled	h	Write-back,	I/O cached,	Disk cache enabled

LSI MegaRAID SAS 1078 with 512 MB Cache, 2.5" 15 krpm Hard Disks

In contrast, the write throughput is more dependent on the cache settings. In order to achieve optimum performance, it is necessary to use the "Disk cache enabled" option as the optimum cache setting. The improvement achieved is roughly 60%.

The importance of optimal cache settings can be seen particularly clearly with RAID 5. The diagram shows that sequential write throughput increases considerably, by a factor of 30, as a result of enabling the controller cache with the option "Write-back" and achieves even higher values than with sequential read, although an additional parity block has to be calculated and written for write. On the other hand, the cache settings have less impact on throughput with sequential read. It is interesting to see how counterproductive the effect of enabling the I/O cache is on read throughput particularly for reads.

The following graphic shows that during random access in RAID 1 the combination of the “Write-back” and “Disk cache enabled” options has a decisive influence on throughput. The improvement in throughput is 48% or 56%, depending on whether access is with 64 KB or 8 KB block size.



	8 KB random 67% reads								64 KB random 67% reads							
	a	b	c	d	e	f	g	h	a	b	c	d	e	f	g	h
RAID 1	2.3	2.7	2.6	2.7	2.8	3.5	2.6	2.7	15.6	17.6	16.4	11.9	19.2	23.1	16.5	11.9
RAID 5	2.6	3.0	2.7	3.0	4.2	4.2	4.1	4.1	15.3	16.4	15.3	16.2	19.6	23.5	22.4	21.9

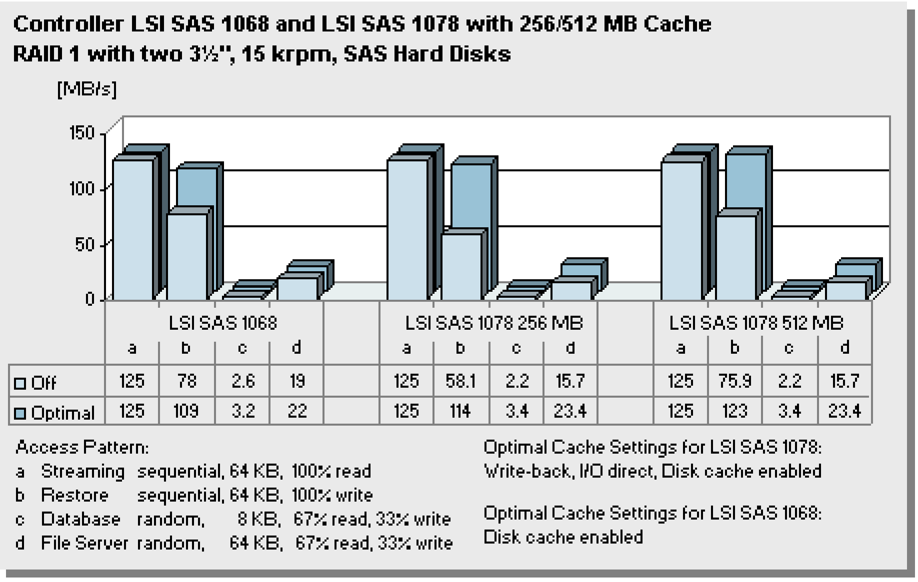
LSI MegaRAID SAS 1078 with 512 MB Cache, 2.5" 15 krpm Hard Disks

For random access in RAID 5 it is advisable to enable the controller cache by setting the write-mode option to “Write-back”. In order to achieve optimal throughput it is also necessary to enable the disk cache. As a result of these optimal cache settings improvements in throughput of 54% or 60% can be achieved, depending on the block size used.

More detailed information about this topic is available in the paper Performance Report - Modular RAID for PRIMERGY.

Controller comparison

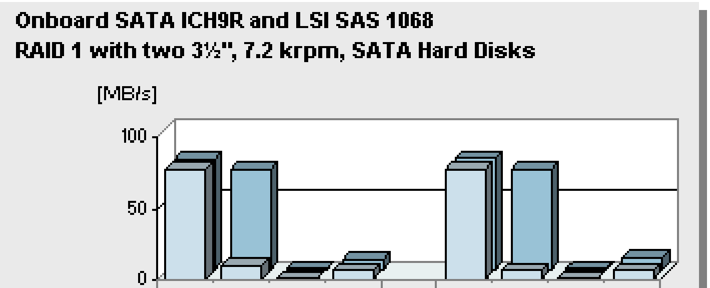
The following comparison depicts the throughputs of the various controllers. The measurements were made with the same hard disk types in the same RAID 1array. The diagram shows the throughputs achieved with disabled caches (Off) and with optimal cache settings (Optimal).



	LSI SAS 1068				LSI SAS 1078 256 MB				LSI SAS 1078 512 MB			
	a	b	c	d	a	b	c	d	a	b	c	d
Off	125	78	2.6	19	125	58.1	2.2	15.7	125	75.9	2.2	15.7
Optimal	125	109	3.2	22	125	114	3.4	23.4	125	123	3.4	23.4

You can see that the cache settings for sequential read access do not have any or only a very minor influence on throughput, regardless of which controller is used. The throughput values achieved are equivalent to the maximum possible values. For sequential write access it is possible to achieve an increase in throughput by optimizing the cache settings. Throughput increases by about 66% with the LSI MegaRAID SAS 1068 controller. The best throughput values were achieved with the LSI MegaRAID SAS 1078 controller with a 512 MB controller cache. The difference in performance to the LSI MegaRAID SAS 1078 controller with a 256 MB controller cache is about 3.2% with optimal cache settings and sequential write access. The difference in performance to the LSI MegaRAID SAS 1068 controller is approximately 2.2%. If the LSI MegaRAID 1078 controller with 256 MB and 512 MB cache is compared with the LSI MegaRAID 1068 controller during random access with 8KB and 64KB, then the throughput difference is between 5 and 6.5%.

When the onboard SATA ICH9R controller is compared with the LSI MegaRAID SAS 1068 controller, it is evident that both controllers offer roughly the same performance when measurements are carried out with the same SATA hard disks and optimum cache settings. The variations in performance are within the range of measuring accuracy.



	Onboard SATA ICH9R					LSI SAS 1068			
	a	b	c	d		a	b	c	d
HD Cache off	77.46	9.85	0.81	6.03		77.14	6.73	0.88	6.30
HD Cache on	77.46	70.26	0.97	7.24		77.97	70.23	1.14	8.57

Access Pattern:
a Streaming sequential, 64 KB, 100% read
b Restore sequential, 64 KB, 100% write
c Database random, 8 KB, 67% read, 33% write
d File Server random, 64 KB, 67% read, 33% write

Relevant differences in throughput were found in sequential writing and deactivated disk cache. In this access pattern, the onboard SATA controller is 44% faster than the LSI MegaRAID SAS 1068 controller, however, the absolute throughput values in both cases are less than 10 MB/s, while the activation of the disk cache on both controllers leads to an increase in performance by a factor of 7 or 10 to 70 MB/s.

During random access with enabled disk cache, on the other hand, the LSI MegaRAID SAS 1068 controller offers about 18% more performance than the onboard SATA controller.

The onboard SATA controller is implemented directly on the motherboard of the server in the Intel ICH9R chipset. The RAID stack is handled by the server CPU. The increased load on the CPU depends on the access pattern and block size. The CPU load increases by up to 5 percentage points in this example, although the CPU load in many small data blocks is generally higher. Although the onboard SATA controller features good performance figures, the benefits of the LSI SAS 1068 controller should not go unmentioned. The advantages of the LSI SAS 1068 controller are not just better performance during random access to the SATA hard disks, but also greater flexibility and scalability. The LSI SAS 1068 controller supports more RAID levels, and 3½" and 2½" SAS hard disks with 10 krpm and 15 krpm.

Thanks to the higher rotational speed (15 krpm) of the SAS hard disks, throughputs increases of between about 55% and 193% can be achieved with all access modes in comparison with the SATA hard disks (7.2 krpm). The user must decide for himself whether his needs are best covered by a less expensive solution with lower performance or a more expensive, higher performance solution.

Conclusion

With the “Modular RAID” concept, the PRIMERGY TX150 S6 offers a plethora of opportunities to meet the various requirements of different application scenarios.

The entry-level controller, represented by the LSI MegaRAID SAS 1068 controller, offers the basic RAID solutions RAID 0, RAID 1 and RAID 1E and supports these RAID levels with a very good performance.

The “high-end” controller, represented by the LSI MegaRAID SAS 1078 controller, offers all today’s current RAID solutions; for the PRIMERGY TX150 S6, which can be expanded with up to eight internal hard disks, this can be RAID levels 0, 1, 5, 6, 10, 50 and 60. This controller is supplied with a 256 MB or 512 MB controller cache and can as an optional extra be secured with a BBU. Various options for setting the use of the cache enable controller performance to be flexibly adapted to suit the RAID levels used.

Use of RAID 5 or RAID 6 enables the existing hard disk capacity to be utilized economically for a good performance. However, we recommend a RAID 10 for optimal performance and security.

When connecting SATA hard disks in a RAID 0, RAID 1 or RAID 10 array, it is best to use the onboard SATA RAID controller. The throughputs are comparable with those of LSI MegaRAID SAS 1068/1078 controllers and the higher CPU load on the SATA controller is easily catered for by modern processors. However, with RAID 5 the missing controller cache of the onboard SATA controller becomes particularly evident during sequential write. If importance is attached to optimal performance, then an LSI MegaRAID SAS 1078 controller should be chosen for SATA hard disks.

The PRIMERGY TX150 S6 offers a choice between SATA and SAS, and for SAS hard disks between 2½" hard disks and 3½" hard disks and also different rotational speeds of 10 krpm or 15 krpm. Depending on the performance required, a decision must be taken as to which hard disk type with which rotational speed is to be used. Hard disks with 15 krpm offer an up to 39% better performance. As a result of using 2½" hard disks it is possible – depending on the RAID level – to achieve higher parallelism through the use of more hard disks in the RAID array.

For maximum performance it is advisable, particularly with SATA hard disks or when using a controller without a controller cache, to enable the hard disk cache. Depending on the disk type used and access pattern, the increase in performance is 13-fold. When the hard disk cache is enabled we recommend the use of a UPS.

Benchmark environment

All the measurements presented here were performed with the hardware and software components listed below.

Component	Details
Server	PRIMERGY TX150 S6
Operating system	Windows Server 2003, Enterprise Edition Version: 5.2.3790 Service Pack 1 Build 3790
File system	NTFS
Measuring tool	Iometer 27.07.2006
Measurement data	Measurement file of 8 GB
Onboard SATA RAID	Product: Intel ICH9R Driver Name: megasr.sys Driver Version: 09.21.0914.2007
Controller LSI MegaRAID SAS 1068	Product: LSI RAID 0/1 SAS 1068, Driver Name: lsi_sas.sys, Driver Version: 1.25.05.00
Controller LSI MegaRAID SAS 1078	Product: LSI RAID 5/6 SAS 1078 Driver name: msas2k3.sys, driver version: 2.17.0.32 Firmware package: 6.0.1-0074, firmware version: 1.11.32-0307 BIOS version: NT10 Controller cache: 256 MB or 512 MB
Hard Disk SATA, 3½", 7.2 krpm	Western Digital WD1600AAJS, 160 GB
Hard Disk SAS, 2½", 10 krpm	Seagate ST973402SS, 73 GB
Hard Disk SAS, 2½", 15 krpm	Seagate ST973451SS, 73 GB
Hard Disk SAS, 3½", 10 krpm	Seagate ST373355SS, 73 GB
Hard Disk SAS, 3½", 15 krpm	Seagate ST373455SS, 73 GB

OLTP-2

Benchmark description

OLTP stands for Online Transaction Processing. The OLTP-2 benchmark is based on the typical application scenario of a database solution. In OLTP-2 database access is simulated and the number of transactions achieved per second (tps) determined as the unit of measurement for the performance of the system measured.

In contrast to benchmarks such as SPECint and TPC-E, which were standardized by independent bodies and for which adherence to the respective rules and regulations are monitored, OLTP-2 is an internal benchmark of Fujitsu Technology Solutions. The partially enormous hardware and time expenditure for standardized benchmarks has been reduced to a reasonable degree in OLTP-2 so that a variety of configurations can be measured within an acceptable period of time.

Even if the two benchmarks OLTP-2 and TPC-E simulate similar application scenarios using the same workload, the results cannot be compared or even treated as equal, as the two benchmarks use different methods to simulate user load. OLTP-2 values are typically similar to TPC-E values. A direct comparison, or even referring to the OLTP-2 result as TPC-E, is not permitted, especially because there is no price-performance calculation.

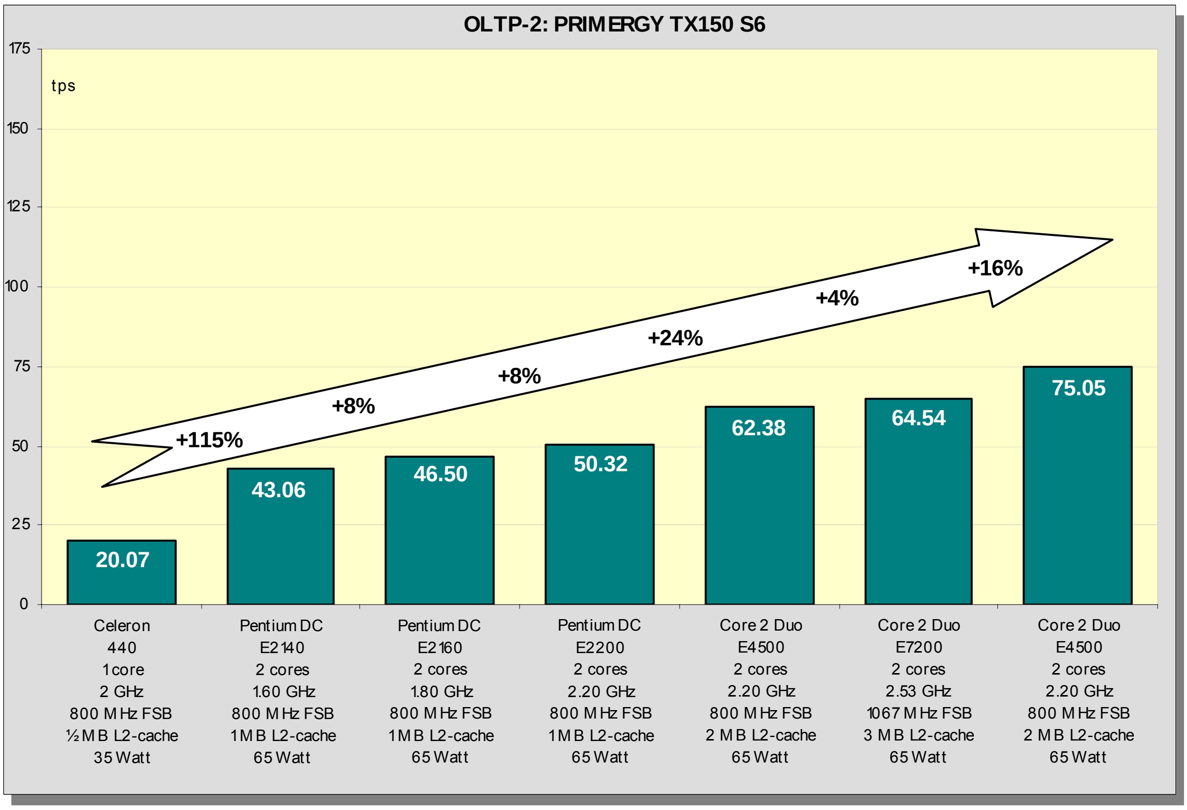
Benchmark results

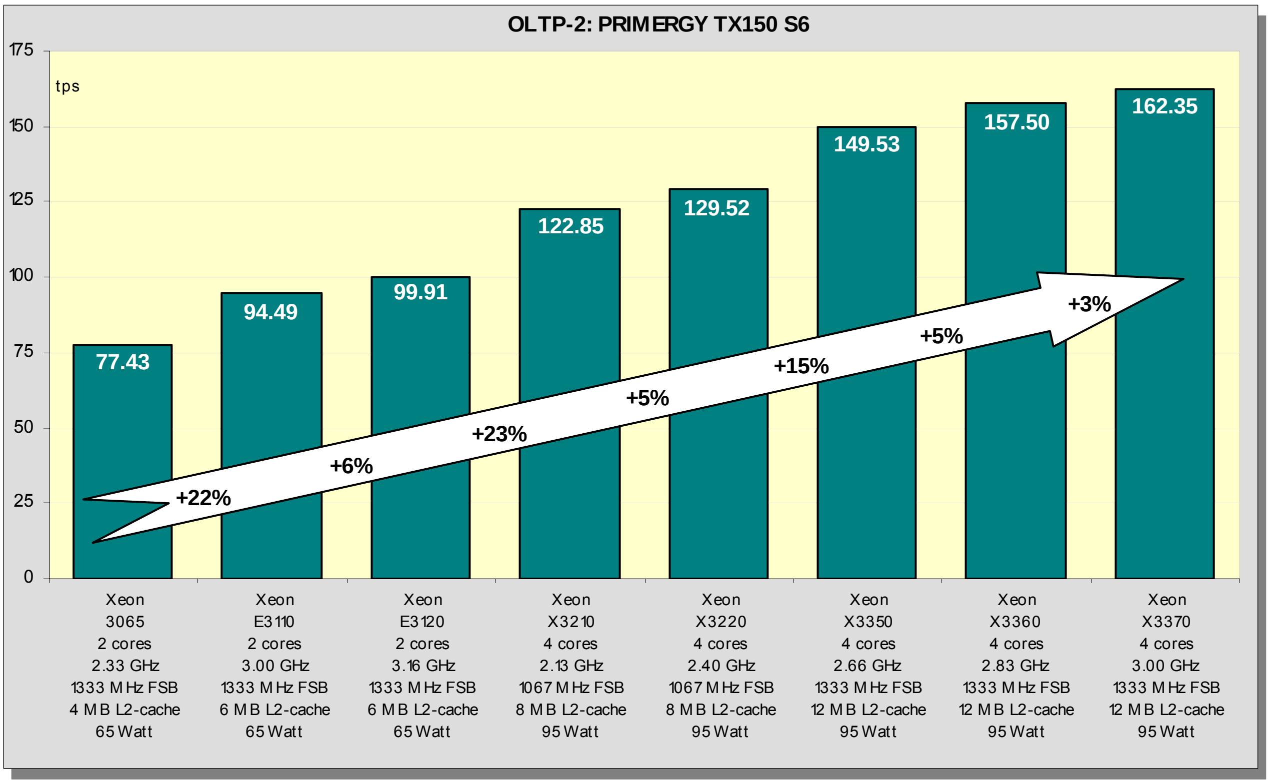
The PRIMERGY TX150 S6 has been released with a large number of processor types. The following table is an overview which processors have been measured with the OLTP-2 benchmark.

Processor	Cores/ Chip	GHz	L2 cache	FSB	TDP	tps
Celeron 440	1	2.00	½ MB per chip	800 MHz	35 watt	20.07
Pentium Dual-Core E2140	2	1.60	1 MB per chip	800 MHz	65 watt	43.06
Pentium Dual-Core E2160	2	1.80	1 MB per chip	800 MHz	65 watt	46.50
Pentium Dual-Core E2200	2	2.20	1 MB per chip	800 MHz	65 watt	50.32
Core 2 Duo E4500	2	2.20	2 MB per chip	800 MHz	65 watt	62.38
Core 2 Duo E4600	2	2.40	2 MB per chip	800 MHz	65 watt	64.54
Core 2 Duo E7200	2	2.53	3 MB per chip	1067 MHz	65 watt	75.05
Xeon 3065	2	2.33	4 MB per chip	1333 MHz	65 watt	77.43
Xeon E3110	2	3.00	6 MB per chip	1333 MHz	65 watt	94.49
Xeon E3120	2	3.16	6 MB per chip	1333 MHz	65 watt	99.91
Xeon X3210	4	2.13	4 MB per 2 cores	1067 MHz	95 watt	122.85
Xeon X3220	4	2.40	4 MB per 2 cores	1067 MHz	95 watt	129.52
Xeon X3350	4	2.67	6 MB per 2 cores	1333 MHz	95 watt	149.53
Xeon X3360	4	2.83	6 MB per 2 cores	1333 MHz	95 watt	157.50
Xeon X3370	4	3.00	6 MB per 2 cores	1333 MHz	95 watt	162.35

All results were determined on the basis of the operating system Microsoft Windows Server 2008 Enterprise x64 Edition and the database SQL Server 2008 Enterprise x64 Edition. OLTP-2 benchmark results depend to a great degree on the configuration options of a system with hard disks and their controllers. Therefore, the system was equipped with two dual-channel Fibre Channel controllers that were connected to a total of 180 hard disks via two FibreCAT CX500. See the Benchmark environment section for further information on the system configuration.

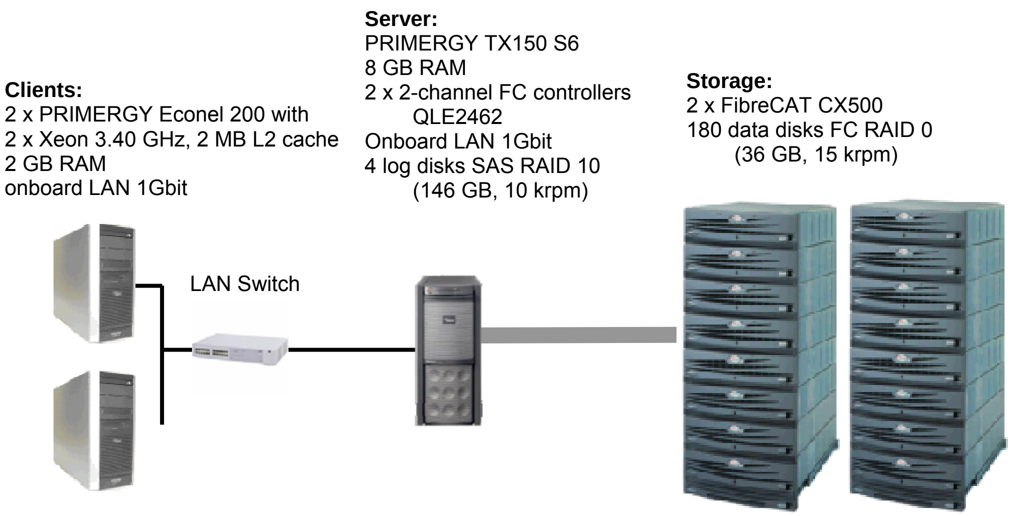
The diagrams below show the OLTP-2 performance data for PRIMERGY TX150 S6 separated in two groups: Celeron, Pentium Dual-Core and Core 2 Duo processors and a second group with Xeon processors. There is a high increase of +115% at Celeron to Pentium Dual-Core related to doubling the number of cores and the L2 cache size. The scaling over all other processor types is about +3% to +25% and depends on processor- and front-side bus frequency increase, larger number of cores and bigger L2 cache.


OLTP-2: PRIMERGY TX150 S6
tps
175
150
125
100
75
50
25
0
20.07
43.06
46.50
50.32
62.38
64.54
75.05
+115%
+8%
+8%
+24%
+4%
+16%
Celeron 440 1 core 2 GHz 800 MHz FSB ½ MB L2-cache 35 Watt
Pentium DC E2140 2 cores 1.60 GHz 800 MHz FSB 1 MB L2-cache 65 Watt
Pentium DC E2160 2 cores 1.80 GHz 800 MHz FSB 1 MB L2-cache 65 Watt
Pentium DC E2200 2 cores 2.20 GHz 800 MHz FSB 1 MB L2-cache 65 Watt
Core 2 Duo E4500 2 cores 2.20 GHz 800 MHz FSB 2 MB L2-cache 65 Watt
Core 2 Duo E7200 2 cores 2.53 GHz 1067 MHz FSB 3 MB L2-cache 65 Watt
Core 2 Duo E4500 2 cores 2.20 GHz 800 MHz FSB 2 MB L2-cache 65 Watt


OLTP-2: PRIMERGY TX150 S6
tps
175
150
125
100
75
50
25
0
77.43
94.49
99.91
122.85
129.52
149.53
157.50
162.35
+22%
+6%
+23%
+5%
+15%
+5%
+3%
Xeon 3065 2 cores 2.33 GHz 1333 MHz FSB 4 MB L2-cache 65 Watt
Xeon E3110 2 cores 3.00 GHz 1333 MHz FSB 6 MB L2-cache 65 Watt
Xeon E3120 2 cores 3.16 GHz 1333 MHz FSB 6 MB L2-cache 65 Watt
Xeon X3210 4 cores 2.13 GHz 1067 MHz FSB 8 MB L2-cache 95 Watt
Xeon X3220 4 cores 2.40 GHz 1067 MHz FSB 8 MB L2-cache 95 Watt
Xeon X3350 4 cores 2.66 GHz 1333 MHz FSB 12 MB L2-cache 95 Watt
Xeon X3360 4 cores 2.83 GHz 1333 MHz FSB 12 MB L2-cache 95 Watt
Xeon X3370 4 cores 3.00 GHz 1333 MHz FSB 12 MB L2-cache 95 Watt

Benchmark environment

Microsoft Windows Server 2008 Enterprise x64 Edition and database SQL Server 2008 Enterprise x64 Edition.



Literature

PRIMERGY Systems	http://ts.fujitsu.com/primergy
PRIMERGY TX150 S6	http://docs.ts.fujitsu.com/dl.aspx?id=af1d8493-dfbb-478e-89c9-92e917929234
PRIMERGY Performance	http://ts.fujitsu.com/products/standard_servers/primergy_bov.html
OLTP-2	Benchmark Overview OLTP-2 http://docs.ts.fujitsu.com/dl.aspx?id=e6f7a4c9-aff6-4598-b199-836053214d3f
SPECcpu2006	http://www.spec.org/osg/cpu2006
	Benchmark Overview SPECcpu2006 http://docs.ts.fujitsu.com/dl.aspx?id=1a427c16-12bf-41b0-9ca3-4cc360ef14ce
SPECjbb2005	http://www.spec.org/jbb2005
	Benchmark Overview SPECjbb2005 http://docs.ts.fujitsu.com/dl.aspx?id=5411e8f9-8c56-4ee9-9b3b-98981ab3e820
SPECpower_ssj2008	http://www.spec.org/power_ssj2008
SPECweb2005	http://www.spec.org/web2005
	Benchmark Overview SPECweb2005 http://docs.ts.fujitsu.com/dl.aspx?id=efbe8db4-7b1b-481e-bdee-66bdfa624b57
StorageBench	Performance Report – Modular RAID for PRIMERGY http://docs.ts.fujitsu.com/dl.aspx?id=8f6d5779-2405-4cdd-8268-1f948ba050e6
	http://www.iometer.org

Contact

PRIMERGY Hardware
PRIMERGY Product Marketing
mailto:Primergy-PM@ts.fujitsu.com

PRIMERGY Performance and Benchmarks
PRIMERGY Performance and Benchmarks
mailto:primergy.benchmark@ts.fujitsu.com

Published by department:

Enterprise Products
PRIMERGY Server
PRIMERGY Performance Lab
mailto:primergy.benchmark@ts.fujitsu.com

Internet:
http://ts.fujitsu.com/primergy
Extranet:
http://partners.ts.fujitsu.com/com/products/servers/primergy